STUDY GUIDE for Molecular Biology

SECOND EDITION

David Freifelder

STUDY GUIDE

for

MOLECULAR BIOLOGY

2nd Edition

David Freifelder
University of California
San Diego

JONES AND BARTLETT PUBLISHERS, INC.
BOSTON PORTOLA VALLEY

Copyright © 1987 by Jones and Bartlett Publishers, Inc. All rights reserved. No part of the material protected by this copyright notice may be reproduced or utilized in any form, electronic or mechanical, including photocopying, recording, or by any information storage or retrieval system, without written permission from the copyright owner.

Editorial offices: Jones and Bartlett Publishers, Inc., 30 Granada Court, Portola Valley, CA 94025.

Sales and customer service offices: Jones and Bartlett Publishers, Inc., 20 Park Plaza, Boston, MA 02116.

Library of Congress Cataloging-in-Publication Data

Freifelder, David Michael, 1935–
Study guide for Molecular biology.

1. Molecular biology—Examinations, questions, etc.
2. Molecular biology—Problems, exercises, etc.
3. Freifelder, David Michael, 1935– . Molecular
biology. I. Freifelder, David Michael, 1935–
Molecular biology. II. Title.
QH506.F73 1986 Suppl. 574.8′8′076 86-20589

ISBN 0-86720-070-7

Printed in the United States of America

10 9 8 7 6 5 4 3

HOW TO USE THE *STUDY GUIDE*

The *Study Guide for Molecular Biology* is designed both as an aid in learning molecular biology and to enable students to determine whether they have succeeded in mastering this highly interdisciplinary subject, which draws so heavily on microbiology, genetics, biochemistry, and cell biology. To achieve this goal, 1292 questions and problems have been provided, each with answers and explanations, when appropriate. Some material that was not included in *Molecular Biology* is also included. The *Guide* is divided into 24 chapters, matched to the chapters in *Molecular Biology*, 2nd Edition. Each is subdivided into nine sections: Chapter Summary, Bold Terms, Drill Questions, Answers to Drill Questions, Problems, Answers to Problems, More Challenging Problems, and Answers to More Challenging Problems. The student should use these parts in the following way:

Chapter Summary. This section summarizes the major concepts and facts in the chapter. The summaries should be read to be sure that each point has been remembered.

Bold Terms. In *Molecular Biology* important terms and definitions are printed in boldface letters when they are first mentioned. These terms are the vocabulary of molecular biology and *must* be understood. This section of the *Study Guide* lists all of these terms. The student should not proceed further unless each term can be defined and understood. If forgotten, the original entry can be found by looking at the lowest page number in the Index of the text.

Additional Information. In this section some of the more complex topics are explained in a slightly different way from what is given in the text. However, for the most part, the section consists of information that (1) was excluded from the text for want of space, or (2) appeared in the current literature after the book had been typeset. It is not essential to read this section, but occasionally this material serves as a basis for a problem. I would hope that at least some students will find this additional material interesting.

Drill Questions. These questions are simple and test the most elementary and basic facts and concepts. The student must be sure that *each* Drill Question can be answered with confidence. If this is not the case, he or she should check the answer, and if it is not clear, a return to the text is called for. Attempting to work the Problems or proceeding to subsequent chapters without such a review will surely be a waste of time.

Problems. These problems test whether the student can make use of the facts and concepts learned in the chapter. Most of the Problems are quite straightforward. I would guess that a typical student will be able to work each one; however, it is certainly not necessary to do so. Try at least ten—for example, every second or third problem, for they are clustered by topic. The set of problems in each chapter covers a large number of subjects, with some duplication of material. Many problems are similar to questions that have arisen in the laboratory, especially when experimental results were being interpreted.

More Challenging Problems. These problems, as the name implies, are more difficult than the Problems. Most will require some thinking beyond the material in the text and occasionally will make use of information gained in later chapters, but which you may have learned in earlier courses. A few of these problems should be tried by everyone.

Answers. Answers are given to all questions and problems. The Drill Questions are occasionally answered without explanation, but all Problems are given complete solutions, except when they are especially simple. I urge you to give each Question and Problem a fair try before consulting the answers.

David Freifelder

June, 1986
San Diego, California

CONTENTS

Chapter 1 Systems and Methods of Molecular Biology 1
Chapter 2 Genetic Analysis in Molecular Biology 8
Chapter 3 Macromolecules 26
Chapter 4 Nucleic Acids 38
Chapter 5 DNA Technology 61
Chapter 6 Physical Structure of Protein Molecules 74
Chapter 7 Macromolecular Interactions and the Structure of Complex Aggregates 87
Chapter 8 The Genetic Material 101
Chapter 9 DNA Replication 109
Chapter 10 Repair 131
Chapter 11 Mutagenesis, Mutations, and Mutants 140
Chapter 12 Transcription 157
Chapter 13 Translation I: The Information Problem 174
Chapter 14 Translation II: The Machinery and Chemical Nature of Protein Synthesis 193
Chapter 15 Regulation of the Activity of Genes and Gene Products in Prokaryotes 206
Chapter 16 Regulation in Eukaryotes 225
Chapter 17 Bacteriophages I: Lytic Phages 241
Chapter 18 Bacteriophages II: The Lysogenic Life Cycle 261
Chapter 19 Plasmids 276
Chapter 20 Genetic Recombination Between Homologous DNA Sequences 286
Chapter 21 Transposable Elements 303
Chapter 22 Recombinant DNA and Genetic Engineering 314
Chapter 23 Eukaryotic Viruses 333
Chapter 24 Tumor Viruses and Oncogenes 344

CHAPTER 1

SYSTEMS AND METHODS OF MOLECULAR BIOLOGY

CHAPTER SUMMARY

Molecular biology is a science whose goal is understanding the molecular basis of biological phenomena. It overlaps biochemistry considerably, differing probably only in the extent to which chemical detail is desired. The advance of molecular biology has tended to follow the development of new physical techniques and the merging of disciplines such as genetics, descriptive biology, and physical chemistry. An important approach in molecular biological research is the isolation of macromolecules, determination of their structure, and evaluation of the relation between structure and function. An assumption that is widely used in developing theories and models is that cells have in the course of evolutionary time become efficient and now behave in an efficient and economical manner; this notion has great predictive value in the study of prokaryotes (organisms in which the genetic material is not enclosed in a membrane), though it is less useful in understanding eukaryotes (organisms whose cells contain discrete membrane-bound nuclei). Classical molecular biology has been confined to a variey of free-living unicellular organisms—mainly, bacteria, yeast, and bacterial viruses. In recent years it has become possible to study eukaryotic cells either grown in culture or isolated from living tissues.

Cells have a wide variety of nutritional requirements, with respect to organic compounds; some are able to grow in media in which the only organic substance is a molecule that serves as a source of carbon atoms for metabolism and for assembly of larger molecules. Other cells require a diversity of substances, such as amino acids and vitamins. A bacterium will multiply on a solid growth medium and form a visible cluster of cells called a colony. The number of bacteria in a sample can be determined by spreading cells on the surface of solid growth medium (plating) and measuring the number of colonies that a known volume of the sample can produce. Nutritional requirements can be identified by plating cells on medium lacking a particular nutrient—lack of growth indicates that a nutrient is required.

Bacteriophages are particles that can multiply within bacterial cells. A

typical phage consists of a nucleic acid molecule encased in a protein shell When a phage adsorbs to a bacterium, phage nucleic acid is injected into the cell. The nucleic acid is used as a source of genetic information for production of numerous progeny phage particles, which are ultimately released when the cell lyses. Phage particles can be counted by plating them with an excess of bacteria. A clear spot, called a plaque, develops against a turbid background of the bacteria.

Yeasts are free-living eukaryotes that have both a haploid and a diploid phase. They multiply by budding rather than by the type of cell division characteristic of bacteria. Yeasts are important organisms in current research since they are eukaryotes and have a eukaryotic life style, like that of animal cells, yet have the advantage that they can be manipulated experimentally with the ease of bacteria. Animal cells can be grown in culture, but since they normally grow in organized tissue, their growth is abnormal. When cells are taken from tissue, they normally multiply for a limited number of generations and then die. A few cells survive and continue to grow and form an established cell line. These cell lines retain many features of the original cells but also exhibit differences.

Two types of chemical processes are carried out in cells, degradative and biosynthetic. Both usually proceed by a series of connected chemical reactions called a metabolic pathway. Most of the individual reactions are catalyzed by enzymes. A reaction can occur spontaneously only when the free energy change is negative. However, many essential chemical conversions have a positive free energy change. When this is the case, molecules, such as ATP, which have a very high negative free energy of hydrolysis, are incorporated into the pathway. The ATP molecule typically enters the pathway by reacting with one of the metabolic intermediates to form a molecule that can undergo a further change in a reaction with a negative free energy of hydrolysis. ATP is the primary energy-storing compound in cells. It is synthesized at various stages of lengthy pathways for degrading organic compounds such as sugars; the most prevalent pathways are glycolysis, the Krebs cycle, and oxidative phosphorylation, the combination of which results in the formation of 38 ATP molecules per glucose molecule.

BOLD TERMS

activated intermediate, adenosine triphosphate, agar, anabolic reaction, ascus, ATP, auxotroph, bacterial colony, bacteriophage, biosynthetic pathway, broth, catabolic reaction, cell density, cell membrane, coat, crude extract, culture, degradative pathway, doubling time, electron transport, established cell line, eukaryote, free energy change, glycolysis, group-transfer reaction, intermediates, Krebs cycle, lag phase, lawn, liquid growth medium, log phase, lysis, lysozyme, macromolecules, metabolic pathway, metabolic regulation, minimal medium, nucleases, nucleus, oocyte, oviduct, oxidative phosphorylation, permeases, petri dish, phage, phage head, phage lysate, phosphate-transfer reaction, plaque, plaque assay, plate, plating, precursors, primary culture,

product, prokaryote, proteolytic, protoplast, prototroph, reticulocytes, soft agar, spheroplast, spores, stationary phase, strain, tail, top agar, transport proteins, variety

DRILL QUESTIONS

1 What is the major difficulty in the interpretation of *in vitro* reactions in crude extracts?

2 Is an experiment in which a radioactive DNA precursor is added to a bacterial culture in order to study DNA synthesis an *in vitro* or an *in vivo* experiment?

3 What is the major difficulty in the interpretation of reactions with purified components?

4 What is the principal tenet of the efficiency argument?

5 What assumption is usually made about the relation between structure and function?

6 What is the defining difference between prokaryotes and eukaryotes?

7 Classify the following cells as either prokaryotes or eukaryotes: bacteria, fungi, algae, yeasts, amoebae, wheat cells, human liver cells.

8 Is the biochemistry of the free-living microbial eukaryotes, such as yeast, more closely related to that of bacteria or the higher eukaryotes (such as plants and animals)?

9 What carbon source yields the highest growth rate of bacteria?

10 What genetic symbol would denote an arginine auxotroph?

11 How would you describe the requirement of a prototroph for organic molecules?

12 In what phase of a bacterial culture are all cells growing at the same rate?

13 If the doubling time of a bacterial culture is 20 min, by what factor does the cell density increase in 2 hr?

14 A bacterium fails to grow on minimal agar but grows if leucine is added. How many different amino acids are required for growth of this bacterium?

15 Define (a) lysozyme and (b) lysis.

16 Define haploid and diploid.

17 What is the major phenotypic difference between cells in an established cell line and cells in tissue?

18 Distinguish a primary culture from an established cell line.

19 What structural feature of ATP gives it a high negative ΔG?

20 By what chemical reaction is ATP synthesized?

21 What pathways result in generation of ATP from glucose, and which pathway is responsible for the greatest number of ATP molecules per molecule of glucose?

22 How is the large negative free energy of hydrolysis of ATP utilized?

ANSWERS TO DRILL QUESTIONS

1 Several different reactions may be occurring simultaneously and these could obscure the reaction of interest.

2 An *in vivo* experiment.

3 All naturally occurring components may not be present.

4 Substances are rarely synthesized or consumed needlessly.

5 No structure exists without a function.

6 In a eukaryote the DNA is enclosed in a nuclear membrane; in a prokaryote the DNA is distributed throughout the cell.

7 Only bacteria are prokaryotes.

8 The higher eukaryotes.

9 Glucose

10 Arg^-.

11 A prototroph requires an organic molecule only as a carbon source.

12 Log phase.

13 2 hr = 6 doubling times. Thus, $2^6 = 64$.

14 Just one, leucine.

15 Lysozyme: an enzyme that solubilizes the cell wall of most bacteria. Lysis: breaking open a cell by any means.

16 Haploid: having only one copy of each chromosome. Diploid: having two copies of each chromosome.

17 Cells of an established cell line will grow indefinitely.

18 A primary culture is a culture consisting of cells that are either taken from an organism or are descendants of one of these cells. The cells are capable of growth for only a limited number of generations. An established cell line is a population of cells derived from a primary culture but which has been altered, presumably genetically, so that it is able to multiply indefinitely.

19 The series of negatively charged oxygen atoms in the phosphate group, which repel one another.

20 $ADP + P_i \rightarrow ATP$.

21 Glycolysis, Krebs cycle, and oxidative phosphorylation (the greatest).

22 A phosphate is transferred to an intermediate molecule (by a reaction with a large negative free energy of hydrolysis, and the intermediate is used in subsequent reactions each of which has a negative free energy. Without hydrolysis of ATP the overall reaction (lacking the activated intermediate) would have a positive free energy.

PROBLEMS

1 A bacterial strain can grow on agar supplemented with arginine (Arg), tryptophan (Trp), and leucine (Leu). It fails to grow on agar containing (I) only Arg and Trp or (II) only Leu and Trp. It will grow if (III) only Arg and Leu are present. What is the genotype of the bacterium with respect to these three amino acids?

2 Suggest two different methods that could be used to determine the number of bacteria per milliliter of culture. Would they give the same value?

3 If you were asked to give an unambiguous definition of a virus, how would you do it?

4 Suggest several reasons why the product of an enzymatic reaction might not be detected in a crude extract?

5 A bacterium divides every 35 minutes. If a culture containing 10^5 cells per ml is grown for 175 minutes, what will be the cell concentration?

6 Suggest reasons for bacteria growing in liquid medium to enter stationary phase.

7 What is the cell density of a bacterial culture if plating 0.1 ml of a 10^5-fold dilution of the culture yields 68 colonies?

8 *E. coli* has a cylindrical shape about 1 mm in diameter and 3 mm long. The doubling time of *E. coli* when growing on nutrient agar is about 25 minutes. After 12 hours of growth, a colony is roughly 2 mm in diameter and 0.5 mm high. Have all of the cells been growing for 12 hours?

9 Figure 1-2 in the text showed a growth curve for a bacterial culture in which cell number, assayed by plating, is plotted against time. A similar curve can be obtained by plotting the optical density of the culture against time because the absorbance is primarily caused by scattering from the cells. It can be shown easily that the absorbance is directly proportional to the dry weight of the culture. Such a plot is not identical to that of Figure 1-2 in that stationary phase, as determined by the absorbance no longer rising, appears to begin about 1/2 generation sooner. Can you explain this difference?

10 10^7 phage are plated in the standard way with the usual number of bacteria to form a lawn (10^8 bacteria). No plaques are seen and the plate appears clear rather than turbid. Explain.

11 When a bud falls off a mother yeast cell, the mother cell is left with a small spot on its cell wall called a bud scar. The daughter cell does not have a scar. Suppose a single daughter cell is allowed to produce a culture consisting of 8 cells. How are the bud scars distributed among the cells?

ANSWERS TO PROBLEMS

1 I shows that Leu is required. II implies that Arg is needed. III indicates that Trp is not needed. Thus, the genotype is *arg*$^-$*leu*$^-$.

2 A direct method is to use a microscope to count the number of bacteria in a known volume of culture. Plating to get cell number is another procedure. Plating might give a smaller number if a fraction of the bacteria is either dead or unable to divide.

3 Two properties define a virus: it is a nucleoprotein particle, in which the nucleic acid is contained in a protein-containing shell, and it can replicate, but only within a living cell.

4 The enzyme might be inactivated when the cells are broken; the concentration of the enzyme or of the reactants might be too low; the product might be unstable; the product might be destroyed by a second reaction; an enzyme inhibitor might be activated.

5 Growth is for 5 generations. Thus, the initial number increases by a factor of 2^5 or 32 and the final concentration is $32 \times 10^5 = 3.2 \times 10^6$ cells/ml.

6 Exhaustion of nutrients; inability of atmospheric oxygen to dissolve in the growth medium as rapidly as it is being consumed; change of pH of the medium. The second item is the most important.

7 6.8×10^7.

8 Yes. The colony contains about 7×10^8 bacteria. If every bacteria had divided every 25 minutes, there would have been about 10^9 cells.

9 By the optical measurement stationary phase would begin when the cells stop synthesizing material and hence stop growing. However, cell division, which is measured by the plating test, can still occur for a brief period after growth has stopped.

10 After one round of replication all bacteria have been infected. Thus, there are no bacteria to form a lawn and hence no plaques can be seen.

11 1, 3 scars; 1, 2 scars; 2, 1 scar; 4, no scars. The original cell has budded three times and has 3 scars. The first daughter cell has budded twice and has 2 scars. Of the remaining 6 cells, 2 are second-generation daughters and have budded once (1 scar each) and 4 are third-generation daughters and have not budded at all.

CHAPTER 2

GENETIC ANALYSIS IN MOLECULAR BIOLOGY

CHAPTER SUMMARY

Genetic analysis has been important in the development of molecular biology, because genetic arguments have often led to testable models of molecular processes and molecular interactions. Genetic arguments can rarely lead to rigorous conclusion about molecules but frequently limit possible explanations. They are usually based on one or more of the following: the relation between the genotype and phenotype of an organism, the number of genes contributing to a phenotype, gene order, and the spacing between genes. The genotype is the genetic make-up of the organism. It is determined by genetic tests for the presence of the wild-type allele or particular mutant alleles. Phenotype refers to an observed property of an organism. Gene number is determined by complementation tests, and gene order and gene spacing are determined by measurements of genetic recombination frequency. A variety of notational systems are used to indicate genotype and phenotype. The one used for bacteria will be found throughout this book and is particularly simple: (1) gene symbols have three letters, or four when several related genes contribute to a single phenotype, (2) phenotype is capitalized and written in roman type, (3) genotype is not capitalized and written in italic type, (4) activity of a gene or presence of a functional gene is denoted by a superscript +, (5) inactivity is denoted by a superscript –, and (6) resistance or sensitivity to any agent or reagent is denoted by -r or -s respectively. Thus, a bacterium unable to utilize the sugar galactose would have the Gal^- phenotype, which would be caused by a defective *galT*, *galK*, or *galE* gene, and a cell resistant to the antibiotic ampicillin would have the Amp-r phenotype, resulting from the presence of the *amp-r* allele. Eukaryotic microorganisms utilize related systems, but often the genotype is indicated in capitalized italic symbols, as in *CYT*.

Mutants, organisms containing one or more mutations, are important tools of molecular genetics. A frequent approach to a problem involves comparing the phenotypes of a wild-type organism and of mutants containing various mutations. Absolute-defective mutations yield a mutant phenotype in all

conditions. Conditional defectives are mutant only in certain conditions, for example, above a certain temperature (temperature-sensitive mutation) or in the presence of particular mutant genes (suppressor mutation). Mutations arise by mutagenesis, and mutants are often selected from populations of organisms. A mutation is a change in the genetic material. If a single change has occurred, the mutant is a point mutant; if two changes have occurred, the mutant is a double mutant. Occasionaly, a portion of a gene or of the genetic material is absent; then, the mutation is said to be a deletion. Organisms with the wild-type phenotype can, at low frequency, be isolated from populations containing a point mutation. The process of producing such an organism is called reversion and the organism is called a revertant. The term reverse mutation is also used to describe the change. The probability of reversion, the reversion frequency, gives information about the nature of the original mutation; for example, double mutations revert at a barely detectable frequency and deletions (except for certain rare classes) do not revert.

Mutants and mutations have been used in the following ways: (1) a mutation often defines a function, (2) a mutation can introduce a biochemical block that can help in elucidating a metabolic pathway, (3) mutations give information about metabolic regulation, (4) mutations enable a biochemical entity to be matched with a biochemical function or an intracellular protein, (5) a mutation can locate the site of action of an environmental agent, (6) a mutation can indicate a relation between apparently unrelated biochemical systems, and (7) certain types of mutations and revertants indicate that two proteins interact.

Genetic recombination is a process in which segments of genetic material become rearranged or reassorted. Genetic recombination events, involving exchange of genetic material between two chromosomes, are useful in molecular biology. For example, organisms carrying novel combinations of mutations can be produced by genetic recombination. Valuable information can also be obtained by measuring the frequency of genetic recombination, that is, the number of recombinants divided by the number of participating parents. Although it is not always strictly true, the recombination frequency between two mutations is often proportional to the physical separation between the mutations in the genetic material. Even in the rare cases when this is not true, recombination frequencies can be used to determine the order of mutations and thereby to define a genetic map. Gene order and gene separation is important in understanding functional relations between genes. The principle of genetic mapping is that genetic distances are additive, if they are neither too large nor too small. Thus, if the recombination frequencies between genes *a* and *b* are 1% and between *b* and *c* are 3%, the frequency between *a* and *c* is $1 + 3 = 4\%$, if the order is *a b c*, or $3 - 1 = 2\%$, if the order is *c a b*. Three–point crosses are valuable in determining gene order and are based on the principle that a combination requiring two exchange events will arise at a frequency that is the product of the frequencies of the individual exchanges.

Complementation is a nonrecombinational method for determining the minimum number of genes that contribute to a particular phenotype. The principle of the technique is the following. If a diploid or partial diploid cell contains two mutations, each of which produces a mutant phenotype, the doubly mutant cell will have the mutant phenotype, if each mutation resides in a copy of the same gene, and will have the wild-type phenotype if the mutations are in different genes. A complementation test assigns a set of mutations to particular classes, each class representing one gene. Thus, a complementation test indicates the minimum number of genes contributing to the phenotype; the number is a minimum because mutations in some genes may not have been isolated.

An important type of genetic system in bacteria is mediated by certain plasmids that enable a donor cell to transfer either the plasmid or the chromosome to a recipient cell. The best-studied system is that mediated by the F, or sex, plasmid in *E. coli*.

Cells containing F can form pairs with cells that do not contain F, and F can be transferred by conjugation from the F^+ cell to the F^- cell. Variants of F, called F', contain chromosomal genes and hence transfer these genes when the plasmid is transferred. F' plasmids are useful to form partial diploids, that is, cells that have two copies of a genetic region. F is able to integrate into the *E. coli* chromosome. When it does so, the cell gains the ability to transfer the entire chromosome; such a cell is called an Hfr. Because of random motion of the cells in liquid growth medium, paired Hfr and F^- cells usually break apart before transfer is complete. In this case, recombination occurs between portions of the transferred DNA and the F^- chromosome. Since a genetic marker will not be detected in a recombinant cell unless it has actually been transferred to the cell, it is possible to determine the time at which a particular marker enters an F^- cell by noting when the marker first appears in recombinants.

This type of mapping, which is called time-of-entry mapping, allows the order of chromosomal genes to be determined. If one makes the probably correct assumption that DNA is transferred at a uniform rate, the relative spacing betwen genes can also be determined. The map obtained by a mating with a particular Hfr strain is linear. However, Hfr strains are known which initiate transfer from different loci in the chromosome and which transfer in both directions. Combining the linear maps obtained with various Hfr strains yields a circular map for the entire chromosome.

BOLD TERMS

absolute defective mutant, allele, cis–dominant mutation, complementation, conditional mutant, conjugation, counterselective marker, deletion, deletion-substitution, double mutant, F, females, genetic recombination, genetic map, genome, genotype, Hfr males, linkage, males, mutagenesis, partial diploid, phenotype, physical map, point mutant, recombination frequency, reversion, reversion frequency, revertant, selected marker, sex plasmid, spontaneous mutagenesis, suppressors, suppressor-sensitive mutants, temperature-sensitive mutant, three–factor cross, Ts, wild–type

ADDITIONAL INFORMATION

The general topic of linkage and recombination is strictly based on logic and probabilities, but nonetheless it is often difficult for the beginner. In this section some of the features that frequently cause trouble are reviewed.

The principle of genetic mapping is straightforward and based on the relation between map distance and the probability of exchange. For fairly short distances mapping is done quite directly. For example, the following genetic distances lead to a simple map: *a–b* 1, *b–c* 2, *a–c* 3. The map is *a–1–b–2–c*, and there is no ambiguity because 1 + 2 = 3. In actuality, the sum is rarely so perfect in a particular experiment because of the statistical nature of genetic exchange. For example, in five different experiments, the observed values of *a–b* might be 1.2, 0.85, 1.16, 0.93, and 0.86, which would average to 1.0. If map distances are larger, they are usually not strictly additive. For example, the segment of the map *d–12–e–14–f* might yield a value of *d–f* of 21 rather than 12 + 14 = 26. The most usual reason for such nonadditivity is the increasing probability of double exchanges with increasing distance. That is, the frequency of recombinants is not directly proportional to the frequency of exchange, because between two markers double exchanges, which do not lead to a recombinant array of markers, go undetected. This problem is discussed in all genetics texts; for the purpose of this review, the message is simply that a value for the map distance between *d* and *f* is obtained more reliably by adding the shorter distances *d–e* and *e–f;* this method is optimal because direct determination of *d–f* will surely underestimate the real value. In fact, values such as 12 and 14 map units may already be sufficiently large that they may too be underestimates of the distances *d–e* and *e–f*. At any rate, adding *d–e* and *e–f* will give a value more nearly correct than the direct determination of *e–f;* furthermore, if map distances for even smaller intervals were available, it would be best to sum these instead. In general, when constructing maps, the shortest intervals should be considered as the most reliable and should be used first. However, a warning is in order: in systems in which very small distances can be measured, for example, bacteriophages, in which a recombination frequency of 0.00001 can be determined, distances are rarely additive for these very short distances. The reason is that one exchange may enhance or inhibit the occurrence of a second crossover in the immediate vicinity.

Two precautions must be taken in calculating recombination frequencies from raw data. One is simply the necessity of counting all recombinations whether they are single exchanges or part of a double-exchange event. Consider a cross *AbD* x *aBd* (gene order *A B D*) yielding the results: 729 *AbD* and 739 *aBd* (parental combinations), 160 *Abd* and 148 *aBD* (single exchange in the *B–D* region), 110 *ABd* and 98 *abD* (single exchange in the A–B region), and 9 *ABD* and 7 *abd* (double exchange). The total number of progeny is 2000. In calculating the map distance between *A* and *B* the single exchanges in *A–B* are not sufficient because each double exchange represents an additional exchange in that region. Therefore, the recombination frequency is (110 + 98 + 9 + 7)/2000 = 11.2%, rather than 10.4%, the value that would result if the double exchanges were neglected. Similarly, the recombination frequency for the region *B–D* is (160 + 148 + 9 + 7)/2000 = 16.2%. Note that the double-exchange class is counted twice, once for each interval.

The second precaution is necessary when studying recombination in viral systems. When performing a cross such as *Ab* x *aB* with viruses, it is not unusual that the parental types and the *AB* recombinant can be detected easily, but that the double mutant *ab* is either not detectable easily or not at all. The problem becomes quite severe with multiply mutant viruses, in which plaque size becomes so small that the plaques may actual not be visible. It cannot be overemphasized that measuring recombination frequency requires that both *AB* and *ab* recombinants be counted. However, because recombination is almost always reciprocal, the number of *ab* doubly mutant recombinants can be taken to be equal to the number of *AB* wild-type recombinants. Thus, recombination frequency can be estimated as 2(number of *AB* recombinants)/total. For example, in a phage cross one might obtain 962 *Ab*, 914 *aB*, and 62 *AB*, for a total number of detected phage of 1938. To calculate the recombination frequency, we add 62 *ab* (that is, a number equal to the number of *AB*) both to the number of recombinants and the total detected phage, so the recombination frequency is 2(62)/(1938 + 62) = 124/2000 = 6.2%.

Genes can clearly be ordered by pairwise mapping and combining map distances. However, as pointed out in the text, a three-factor cross enables three genes to be mapped in a single cross. Although this is adequately described in the text, it is worthwhile to remind the student of the essential features of this procedure. Two types of information are obtained by analysis of the data from three-factor crosses—the gene order and the distances—and these types are obtained from different components of the data. The gene order is obtained simply from identifying the two parental combinations of genes (the largest numbers) and the double-exchange classes (the smallest numbers). In the double-exchange recombinants the central markers are exchanged with respect to the parental combinations, so gene order can be determined immediately. Recombination values for the single-exchange classes are not needed for ordering the genes. However, these values provide the recombination frequencies, and hence the map distances, for the intervals in which the exchange has occurred.

DRILL QUESTIONS

1 Does the symbol lac^+ refer to the genotype or the phenotype of a bacterium?

2 What notation would indicate that a bacterium is resistant to penicillin?

3 If a cell has three genes involved in synthesis of proline, namely, *proA*, *proB* and *proC*, would there be any difference in the genotype of a cell designated $proA^+proB^+proC^+$ and one designated pro^+?

4 Referring to Drill Question 3, how would you write the phenotype of a cell whose genotype is $proA^+proB^-proC^+$.

5 How does an absolute defective mutant differ from a conditional mutant?

6 Is a temperature-sensitive mutant an absolute defective or a conditional mutant?

7 The frequency of producing mutant arb^- is 2×10^{-6} per generation and mutant snd^- is 8×10^{-5}. What is the frequency of producing arb^-snd^- in a single event?

8 Rank the following mutations in order of reversion frequency: deletion, point mutation, double mutation.

9 A cell carrying v^-(Ts), a temperature-sensitive mutation in gene *v*, has lost the ability to synthesize substance V. An enzyme Vase, which is involved in synthesizing V, is isolated from a mutant containing v^-(Ts) and shows normal activity over a wide range of temperatures. Is Vase the product of gene *v*?

10 A mutation prevents the synthesis of substance Z and results in the accumulation of large amounts of substance R, which is normally present in only very small amounts. What can probably be said about the relation between R and Z and about the gene product inactivated by the mutation?

11 A linkage group refers to a collection of genes that map in a single unit in the sense that the genes are separable only by genetic recombination; that is, the genes do not freely assort in the Mendelian sense. If all the chromosomes in an organism were mapped, how many linkage groups would be found in (1) a haploid organism with 17 chromosomes per somatic cell, (2) a virus with only one chromosome (which is typical for viruses), and (3) a rat, with 42 chromosomes per somatic cell?

12 Two phages with genotypes x^+y^- and x^-y^+ are crossed. What are the genotypes of the possible recombinant types? Are any of them wild-type?

13 The recombination frequencies between three genes are: *a–b*, 2.6%, *b–d*, 1.4%, and *a–d*, 1.2%. What is the gene order?

14 The recombination frequency between *q* and *r* is 3.7%. Does it matter whether the map is represented *q*–3.7–*r* or *r*–3.7–*q*?

15 In a phage cross *ABD* x *abd*, what are the genotypes of the double-exchange classes, if the map order is *A D B*?

16 Mutants that fail to synthesize a substance X have been found in four complementation groups, none of which are cis-acting. How many proteins are required to synthesize X?

17 In a haploid organism mutation x^- eliminates production of a pigment; another mutation y^- does the same. A partial diploid x^-/y^- cell makes pigment. Do *x* and *y* complement? How many genes are required for pigment formation?

18 Does a cis-dominant mutation define a gene with a gene product?

19 Two phage types infect a single cell. Each phage type contains one of three mutations *1*, *2*, or *3*, which prevent phage production. Phage are produced in the combinations *1–2* and *1–3*, but not in the combination *2–3*. What are the complementation groups?

20 Does the F plasmid make a cell a donor or a recipient?

21 How does an Hfr cell differ from an F^+ cell?

22 Is there only one or several possible sites of integration of F in the chromosome?

23 What is the function of a counterselective marker in an Hfr x F^- mating?

24 Distinguish F^+ and Hfr transfer with respect to the amount of genetic material (DNA) transferred and the intactness of the transferred unit.

25 In a time-of-entry experiment how are recombination frequencies for a particular gene at various times used to determine the time of entry of that gene?

26 In a set of time-of-entry curves all curves do not have a plateau at the same recombination value and the plateau value decreases with increasing time of entry. What is the major reason for this phenomenon?

27 Genes *p*, *f*, and *q* have times of entry 7, 11, and 19 minutes, respectively. What is the gene order and the map distances, in min, from the transfer origin?

28 How are F' plasmids produced?

29 How are F' plasmids used to produce a partial diploid?

ANSWERS TO DRILL QUESTIONS

1 Genotype.

2 Pen-r.

3 No.

4 Pro^-. Note that it is capitalized and roman.

5 An absolute defective mutant has a mutant phenotype in all conditions; a conditional mutant has a mutant phenotype only in certain conditions.

6 Conditional.

7 $(2 \times 10^{-6})(8 \times 10^{-5}) = 1.6 \times 10^{-10}$.

8 The reversion frequency for the point mutation is the highest; the frequency for the deletion is, for all practical purposes, zero.

9 No. If Vase were the product of gene *v*, it would be temperature-sensitive.

10 R is a precursor to Z, perhaps occurring immediately before Z in the biosynthetic pathway. The mutation inactivates an enzyme that causes conversion of R to the next step in the pathway.

11 (1) 17; (2) 1; (3) there is one linkage group per chromosome type and rats are diploid (have two copies of each chromosomes), so the number of linkage groups is (1/2)42 = 21.

12 x^+y^+ (wild-type) and x^-y^-(double mutant).

13 *a d b*. If *a* and *d* were on opposite sides of *b*, *a–d* would be 4%. Thus, the order must be *a d b* or *d a b*. For the first order, *a–d* would be 1.2%; for the second order, *b–d* would have to be larger than *a–b*, which is not the case.

14 No. Unless a third gene has been given as a reference, the order of two genes is written in an arbitrary direction.

15 Remember that the central marker will be exchanged. Thus, the genotypes are *ABd* and *abD*.

16 In general, at least four.

17 The mutations complement. At least two genes are required.

18 No. It usually indicates a regulatory site.

19 Since *1* and *2* complement, they are in different groups. Similarly, *1* and *3* are

in different groups. However, *2* and *3* fail to complement, so they are in the same group. Thus, there are two groups: (*1*) and (*2*, *3*).

20 A donor.

21 In an Hfr cell F is integrated into the chromosome.

22 Many. Each integration site produces a particular Hfr cell line.

23 It prevents the Hfr from producing a colony.

24 An Hfr cell can transfer more DNA, but usually only a fragment of the chromosome is transferred. An F^+ or F' cell generally transfers the intact plasmid.

25 The values are extrapolated back to the time axis; the time corresponding to a frequency of 0% is the time of entry.

26 Hfr and F^- cells are broken apart before chromosomal transfer is complete; thus, a gene transferred later will be present in the recipient at a lower frequency than a gene transferred earlier.

27 The order is *p f q* and the map distances are origin–7–*p*–4–*f*–8–*q*.

28 Aberrant excision of F from an Hfr cell, such that adjacent bacterial genes are contained in the circular plasmid.

29 An F' containing genes *x*, *y*, ... is transferred to a recipient that also contains *x*, *y*, ...

PROBLEMS

1 Approximately 10^8 *E. coli* cells of an auxotrophic strain (the requirements to be determined) are plated on complete medium and form a bacterial lawn. Then a circular piece of velvet cloth is pressed on the surface of the plate. Bacteria from all over the plate stick to the hairs of the velvet. The cloth is then pressed onto the surface of three plates (replica plates) containing solid minimal medium supplemented by either leucine, thymine, or proline. About 10^7 cells are transferred onto each plate from the velvet. The cloth is kept in the same orientation so that a cell that could grow on two different plates would form a colony in the same position on both plates. (This technique is called replica plating.) The plates are then incubated to allow colony formation. No colonies are found on the thymine plate, and one colony each is found on the leucine and proline plates; these colonies are not in the same position. (a) What is the genotype of the strain? (Note that there is not a lawn on any of the supplemented plates.) (b) Explain the colonies that appear on the replica plates.

2 Often dyes are incorporated into agar in order to determine whether a bacterium can utilize a particular sugar as a carbon source. For instance, on EMB agar containing a sugar X, suppose that a bacterium having the phenotype X^+ makes a purple colony and another having the phenotype X^- makes a pink colony. (Fermentation of the sugar produces acid which changes the dyes to yield a purple color.) If the EMB agar contained rhamnose, xylose, and galactose, what would be the color of the colonies produced by bacteria having the following phenotypes? $Rha^+Xyl^+Gal^+$, $Rha^-Xyl^-Gal^+$, $Rha^+Xyl^-Gal^+$, $Rha^-Xyl^+Gal^-$, $Rha^-Xyl^-Gal^-$.

3 A colleague suggests to you that glucose is a substance that freely diffuses through the *E. coli* cell wall. You have in your refrigerator an *E. coli* mutant with the following property. If glucose is the sole carbon source in the growth medium, it grows normally at 34 C but cannot grow at 42 C; with any other carbon source it grows well at both temperatures. (a) Can you conclude that there is a glucose transport system? (b) Suppose, by biochemical techniques, you make the following observation. Glucose is not metabolized in this mutant at 42 C; however, if the cell culture is treated with the enzyme lysozyme, which removes a portion of the cell wall, glucose metabolism can occur at 42 C. Does this change your conclusion?

4 Study of a temperature-sensitive mutant in the gene *memA* indicates that there is no active MemA enzyme at 42 C, although at 34 C active enzyme is present. The enzyme is purified from cells grown at 34 C and its activity is tested at 42 C. It remains active at this temperature. What conclusion can be drawn about the relation between the *memA* gene and the MemA enzyme?

5 In phage BX4 some mutations in the phage gene *G* are compensated for by an additional mutation in gene *H* and some mutations in gene *H* are compensated for by mutations in gene *G*. What do these facts indicate?

6 Given the recombination frequencies between genes *a* and *b*, *a* and *c*, and *c* and *d*, how many more values are needed to construct a map?

7 (a) Suppose the following recombination frequencies occur between the indicated markers: *a* x *c* , 2%; *b* x *c* , 13%; *b* x *d*, 4%, *a* x *b*, 15%, *c* x *d*, 17%; *a* x *d*, 19%. What is the gene order? (b) In the cross *aBd* x *AbD*, what is the frequency of getting *ABD* progeny?

8 A phage with genotype *Abd* is crossed with another phage with genotype *aBD*. The following recombinants are observed at the indicated frequencies: *abd*,

2%; *ABd*, 3%; *aBd*, 0.06%. What is the gene order?

9 In the cross *abcd* x *ABCD*, the following recombinants were found at the indicated frequencies: *ABCd*, 3%; *abcD*, 3%; *AbcD*, 0.03%; *AbCd*, 0.0006%; *ABcD*, 0.06%. What is the gene order?

10 Four genes have the order *a b d e*. The recombination frequencies between various pairs are: *a* x *b*, 1%; *b* x *d*, 2%; *d* x *e*, 3%. (a) What is the frequency of production of *AE* recombinants in the cross *AbDe* x *aBdE*? (b) What fraction of the *AE* recombinants will be *ABDE*?

11 In a cross between two phage having genotypes *EFG* and *efg*, 1000 progeny were analyzed. The number of phage having each of the eight possible genotypes were as follows: *efg*, 396; *EFG*, 406; *eFg*, 23; *efG*, 1; *EfG*, 25; *Efg*, 75; *eFG*, 73; *EFg*, 1. Construct a map showing the positions of the genetic markers.

12 In a cross *ABD*/*abd* x *abd*/*abd* between two diploid individuals the most common progeny are *ABD*/*abd* and *abd*/*abd*, and the least frequent are *aBD*/*abd* and *Abd*/*abd*. What is the gene order?

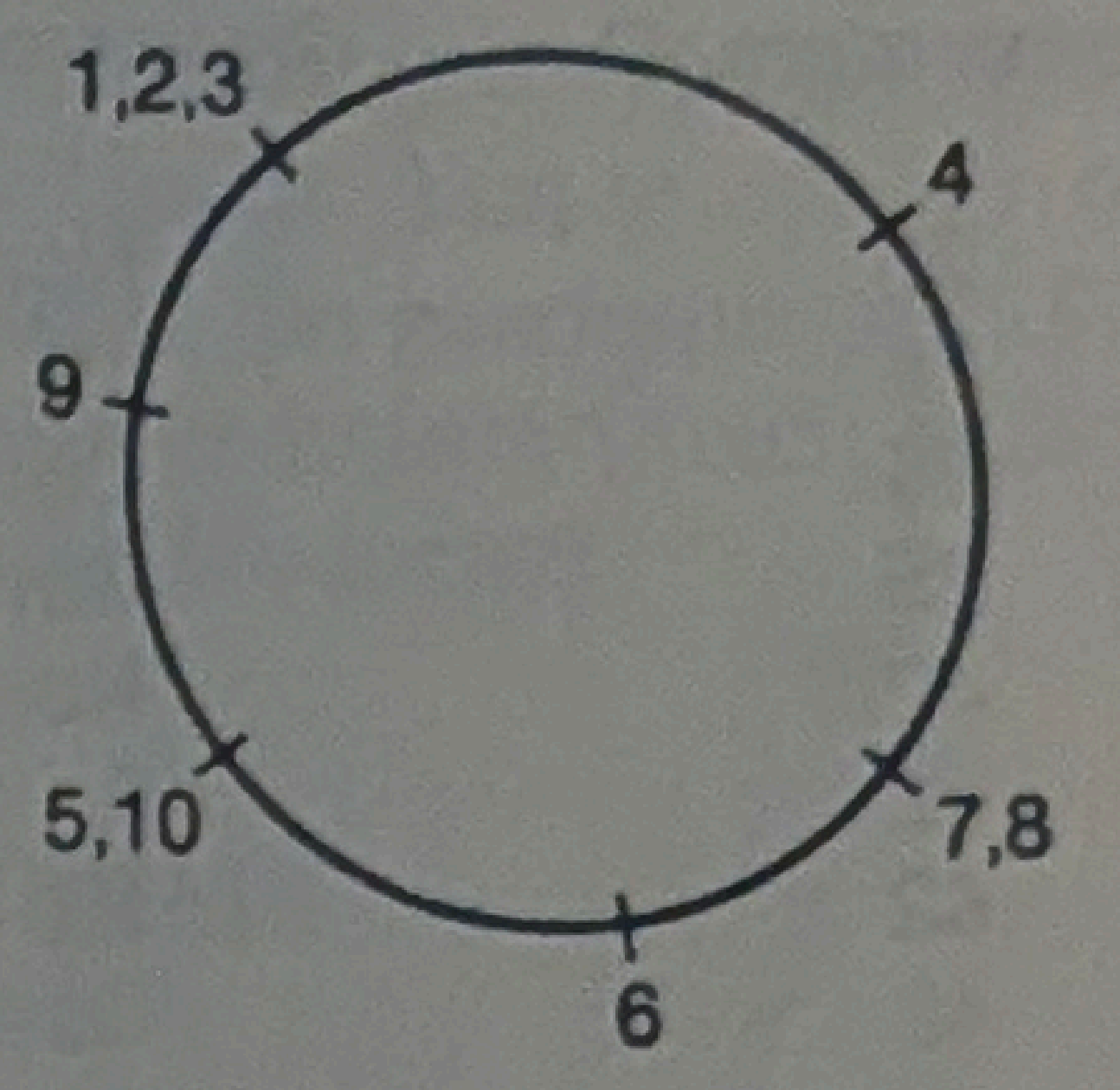
13 A three-factor phage cross *ABD* x *abd* is carried out which yields the fact that the recombinant *AbD* is formed at a much higher frequency than *Abd*. What is the gene order, assuming equal spacing of the genes?

14 A mutation j^-(Ts) eliminates the ability of a cell to synthesize substance J above 40 C. The enzyme Jase, which is needed to synthesize J, is isolated from the mutant and found to be functional above 40 C. What is the minimum number of proteins needed to synthesize J?

15 Do you think that a cis-dominant mutation will always fail to complement all other mutations in a set of related mutations that define a particular metabolic pathway?

16 Four genes *kyuA*, *kyuB*, *kyuC*, and *kyuQ* are known to be required to synthesize substance Q and each biochemical reaction can be detected. The reaction sequence is P→B→A→Q in which the product of a gene *kyuX* is needed to synthesize substance X. Addition of ^{14}C–P yields ^{14}C–Q. (a) A mutant is found for which addition of ^{14}C–P yields ^{14}C–A but no ^{14}C–Q. In what gene is this mutation? (b) Another mutant is found for which there is no conversion of ^{14}C–P to any other substance. Furthermore, addition of ^{14}C–A fails to yield ^{14}C–Q. What kind of mutant is this one?

17 You are interested in the biosynthetic pathway of compound X and isolate 10 different (independently isolated) mutants of E. coli which require compound X for growth. The mutations are mapped and their approximate positions are given in the figure below. (a) What is the minimum number of genes involved in the synthesis of compound X? (b) Why must your answer be the minimum estimate?



18 The complementation data shown in the table below are observed. The numbers refer to particular mutations. The symbols + and – indicate that the two mutations do and do not complement respectively. How many genes are represented? Assign the mutations to the genes.

	Mutants						
	1	*2*	*3*	*4*	*5*	*6*	*7*
1	–	+	+	+	+	+	–
2		–	+	–	+	+	+
3			–	+	+	–	+
4				–	+	+	+
5					–	+	+
6						–	+
7							–

19 Suppose you are given a mutant strain of *E. coli*. In checking its genotype you test growth requirements at two temperatures. You plate 10^8 cells at either 25 C or 42 C on minimal agar plates supplemented with various amino acids. The amount of growth is indicated in the table below.

Supplement	*Temperature, C*	*Growth of colonies*
His, Trp	25	None
His, Leu	25	10
Leu, Trp	25	Confluent
His, Leu, Trp	25	Confluent
His, Trp	42	None
His, Leu	42	8
Leu, Trp	42	12
His, Leu, Trp	42	Confluent

(a) What is the genotype of the strain? (b) Why is there confluent growth (growth covering the entire plate) on the plate containing leucine and tryptophan at 25 C but not at 42 C? (c) Why are there no colonies on the plates containing histidine and tryptophan? (d) What is the expected genotype of the colonies that grow at 42 C on the agar containing histidine and leucine? (e) What is the expected genotype of colonies that grow at 42 C on the agar containing leucine and tryptophan?

20 An Hfr strain with genotype $met^{-}his^{+}leu^{+}trp^{+}$ and which transfer the *met* gene very late is mated with a $leu^{-}met^{+}trp^{-}his^{-}$(Ts) female. The his^{-}(Ts) mutation introduces a requirement for histidine at 42 C. After mating for several hours the mixture is diluted and plated on media having four different supplements. The plates are incubated at 42 C. The supplements in the plates and the number of colonies per plate are the following: His + Trp, 250; His + Leu, 50; Leu + Trp, 500; His, 10. (a) What is the purpose of the met^{-} mutation in the Hfr strain in this experiment? (b) Which genes entered first, second, and third? (c) You now know the relative order of these three markers, but you know nothing about their exact location on the chromosome. What type of experiment would tell you where these markers are located on the chromosome?

21 An Hfr donor whose genotype is $a^{+}b^{+}c^{+}str\text{-}s$ mates with a female whose genotype is $a^{-}b^{-}c^{-}str\text{-}r$; the order of transfer is *a b c*. None of these genes are transferred early, the distance between *a* and *b* is the same as the distance between *b* and *c*, and none of the markers are near *str*. Recombinants are selected as usual by plating on agar lacking particular nutrients and containing streptomycin. Which of the following are true (several answers are)? Explain. (a) $a^{+}str\text{-}r$ colonies > $c^{+}str\text{-}r$ colonies. (b) $b^{+}str\text{-}r$ colonies < $c^{+}str\text{-}r$ colonies. (c) $a^{+}b^{+}str\text{-}r$ colonies < $b^{+}str\text{-}r$ colonies. (d) $a^{+}b^{+}str\text{-}r$ colonies = $b^{+}str\text{-}r$ colonies. (e) Most $a^{+}c^{+}str\text{-}r$ colonies will also be b^{+}. (f) Most $b^{-}c^{+}str\text{-}r$ colonies will also be a^{-}. (g) $a^{+}b^{+}c^{-}str\text{-}r$ colonies < $a^{+}b^{-}c^{-}str\text{-}r$ colonies.

22 Suppose you collect a large number of galactose–requiring (Gal^{-}) bacterial mutants and identify three closely–linked genes (designated *galA*, *galB*, and *galC*) by complementation and rough mapping studies. You wish to order these genes and learn something about the genetic structure of the "galactose region" of DNA. To order the genes, you mate an Hfr having the genotype $bio^{+}gal^{+}str\text{-}s$ with various F^{-}strains having the genotype $gal^{-}bio^{-}str\text{-}r$. The Hfr transfers the *bio* locus later than the *gal* locus. You select $bio^{+}str\text{-}r$ recombinants, measure the fraction of these that have the genotype gal^{+}. For each of the mutations the following fractions are observed: $galA^{-}$, 0.65; $galB^{-}$, 0.72; $galC^{-}$, 0.84. What is the gene order relative to the *bio* locus?

ANSWERS TO PROBLEMS

1 (a) $Leu^{-}pro^{-}$. (b) Since a lawn does not appear on any of the plates, each of which have a single supplement, the auxotroph must have two or more nutritional requirements. A single colony on a plate must be a revertant whose remaining requirements are satisfied by the supplement in that plate. Since each plate contains only one supplement, the auxotroph must require two substances. Hence, colonies on the Leu and Pro plates are Pro^{+} and Leu^{+} revertants, respectively. Thymine is not required by the auxotrophic strain since if it was, growth on the Leu and Pro plates would require reversion of a thy^{-} mutation also, and double reversion is far too unlikely. No colonies are found on the Thy plate because such a colony would have to be a double revertant for the *pro* and *leu* mutations.

2 All would be purple except for the triple mutant.

3 (a) No, because the defect might be in an enzyme required for glucose metabolism rather than for transport. (b) Yes, because after the lysozyme treatment, the cell must be permeable and allow glucose to enter the cell. The fact that making the cell permeable allows glucose metabolism to occur clearly shows that there is a permeability barrier at 42 C, which must be genetically determined.

4 The enzyme is probably not the product of the gene.

5 That the *G*-gene and *H*-gene products probably join together to form an active macromolecular complex. The regions in which mutations occur (which can be compensated or are compensatory) are probably in the binding sites.

6 Two more: for example, one to locate *c* to the right or left of *a* and another to locate *d* to the right or left of *c*.

7 (a) The gene order is *a c b d*. (b) 0.15 x 0.04 = 0.006.

8 *bda*

9 The three classes of frequencies represent single, double, and triple exchanges. If *ABCd* and *abcD* are the products of a single exchange, the order must be *AB CD*.

10 (a) In a cross *AbDe* x *aBdE*, *AE* can result from a exchange anywhere between *A* and *E*. Since the map distance from *A* to *E* is 0.01 + 0.02 + 0.03 = 0.06, *AE* will occur at a frequency of 0.06. (b) To get *ABDE*, one needs exchanges in the intervals *A–B*, *B–D*, and *D–E*; the probability of getting all three exchanges is 0.01 x 0.02 x 0.03= 0.000006. Therefore, the fraction of *AE* that will be *ABDE* is 0.000006/0.06 = 0.0001.

11 *e g f* is the gene order. The spacing between *e* and *g* is three times that between *f* and *g*.

12 The most frequent classes indicate the noncrossover classes. The least frequent are the double crossovers, in which the central markers are exchanged. Thus, the central markers are *A* and *a*, so the gene order is *B A D*. Note that this is the same order as *D A B*; left and right are completely arbitrary in a genetic map.

13 This can be worked out most easily by writing down the cross with the three possible gene orders: (I) *A B D*, (II) *A D B*, and (III) *B A D*. Recalling that single crossovers are much more probable than double crossovers, we can predict that for I *bAD* occurs much less frequently than *Abd*, for II *AbD* occurs with comparable frequency with *Abd*, and that for III *AbD* occurs much more frequently than *Abd*. Hence, order III is the only one that agrees with the data.

14 Two. One is Jase, and the other is the product of the gene in which the j^{-}(Ts) mutation resides.

15 No. It might only regulate the activity of one or a small number of genes. Usually, a cis-dominant mutation fails to complement all mutations that are located in adjacent genes. If a gene in the pathway is separated from the gene cluster, the cis-dominant mutation might complement a mutation in this gene. This phenomenon will become clear when bacterial operons are described in Chapter 15.

16 (a) *kyuQ*. (b) A regulatory mutant that prevents synthesis of each gene product.

17 (a) Six genes. (b) Some of the closely linked mutations might lie in closely linked genes, that is, they might complement one another. Moreover, other genes in the pathway for which no mutants have yet been isolated might exist.

18 Four genes. The groups are: (*1*, *7*), (*2*, *4*), (*5*), and (*3*, *6*).

19 (a) The genotype is his^{-}(Ts)$trp^{-}leu^{-}$. (b) The strain is his^{-}(Ts) and does not require histidine at 25 C. (c) The leucine requirement results from a deletion or double mutation. (d) The genotype is his^{-}(Ts)$leu^{-}trp^{+}$. (e) The genotype is $leu^{-}trp^{-}his^{+}$.

20 (a) Met^{-} is the counterselective marker, and prevents growth of the Hfr cells. (b) *his*, *leu*, *trp*, in that order. (c) An interrupted mating, that is, a time-of-entry experiment.

21 (a) True, because *a* enters before *c*. (b) False, because *b* enters before *c*. (c) True, because some exchanges will be between *a* and *b*. (d) False. See (c). (e) True, because *b* is between *a* and *c*. (f) True, because the a^{-} and b^{-} alleles are near, and on, the same DNA molecule. (g) True, because *b* enters after *a*.

22 The larger the fraction the nearer the particular *gal* mutation is to the *bio* locus. Thus, the gene order is *galA galB galC bio*.

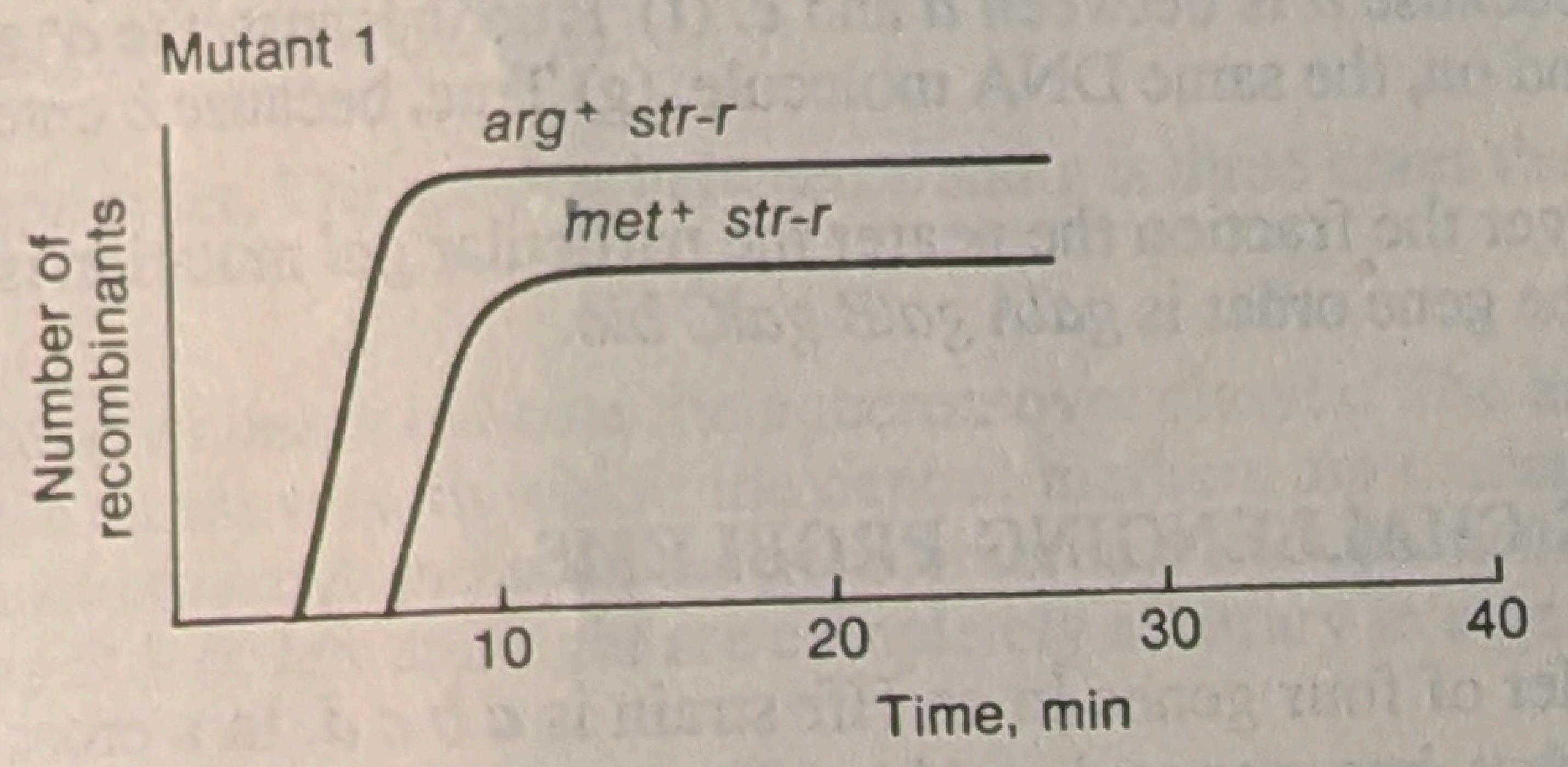
MORE CHALLENGING PROBLEMS

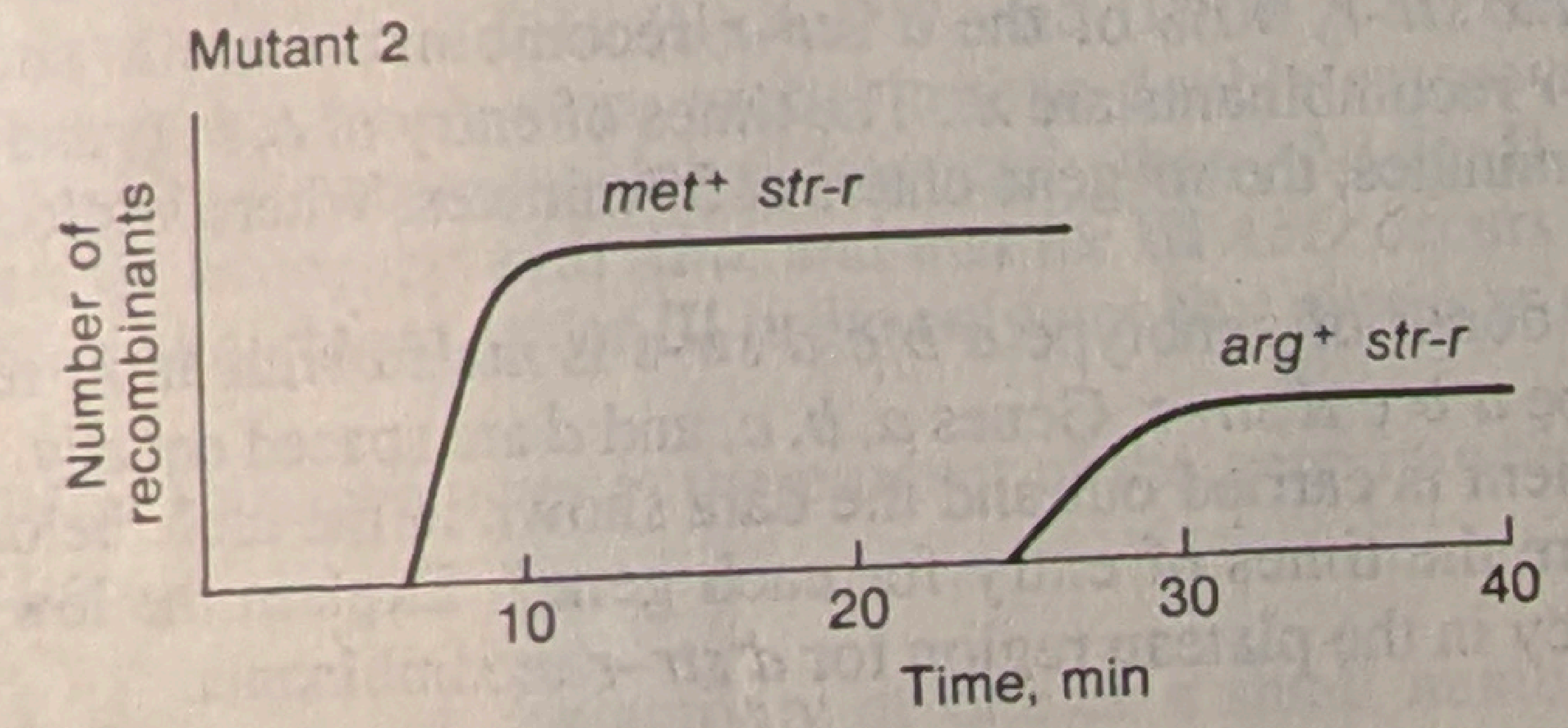
1 The order of four genes in an Hfr strain is *a b c d*. In a cross between an Hfr donor that has genotype $a^{+}b^{+}c^{+}d^{+}x^{-}str\text{-}s$ and a female that has genotype $a^{-}b^{-}c^{-}d^{-}x^{+}str\text{-}r$, 90% of the $d^{+}str\text{-}r$ recombinants are x^{-} and 100% of the $c^{+}d^{+}str\text{-}r$ recombinants are x^{-}. The times of entry of *a*, *b*, *c*, and *d* are 5, 10, 15, and 20 minutes; the *str* gene enters at 55 minutes. Where is *x* located?

2 An Hfr donor of genotype $a^{+}b^{+}c^{+}d^{+}str\text{-}s$ is mated with an F^{-} recipient having genotype $a^{-}b^{-}c^{-}d^{-}str\text{–}r$. Genes *a*, *b*, *c*, and *d* are spaced equally. A time-of-entry experiment is carried out and the data shown in the table below are obtained. What are the times of entry for each gene? Explain the low recombination frequency in the plateau region for $d^{+}str\text{–}r$ recombinants.

Time of mating, in min	*Number of recombinants of indicated genotype per 1000 Hfr*			
	$a^{+}str\text{-}r$	$b^{+}str\text{-}r$	$c^{+}str\text{-}r$	$d^{+}str\text{-}r$
0	0.01	0.006	0.008	0.0001
10	5	0.1	0.01	0.0004
15	50	3	0.1	0.001
20	100	35	2	0.001
25	105	80	20	0.1
30	110	82	43	0.2
40	105	80	40	0.3
50	105	80	40	0.4
60	105	81	42	0.4
70	103	80	41	0.4

3 Suppose you have isolated two independent arginine-requiring (*arg*$^-$) mutant strains from a parent *E. coli* strain which already requires methionine (*met*$^-$) and is resistant to streptomycin (*str-r*). You mate the two mutants (*1* and *2*) with an Hfr strain whose genotype is *arg*$^+$*met*$^+$*str-s*. Using the interrupted mating technique you obtain the time–of entry curves shown in the figure below. Draw a genetic map which explains the difference observed in the two matings.


Mutant 1


Mutant 2

4 After a brief mating between an Hfr whose genotype is *pro*$^+$*pur*$^+$*lac*$^+$ and a female whose genotype is F$^-$*pro*$^-$*pur*$^-$*lac*$^-$*str-r*, many *lac*$^+$*pur*$^+$*pro*$^-$*str-r* recombinants are found. A few *pro*$^+$*lac*$^-$*pur*$^-$*str-r* recombinants also arise, and all of these are Hfr donors. Explain this result and state the location of F in the Hfr. (The three genes *pro*, *lac* and *pur* are very near one another.)

ANSWERS TO MORE CHALLENGING PROBLEMS

1 Between *c* and *d*, because *x* is closely linked to both genes.

2 *a*, 10; *b*, 15; *c*, 20; *d*, 30. Gene *d* is probably very near the *str* locus.

3 The *arg*$^-$ mutants have different map locations and must be in different genes; one is transferred before *met;* the other is transferred after *met*.

4 *Pro* is a terminal marker. F is closely linked to *pro* (that is, all *pro*$^+$ recombinants are donors and hence F$^+$) and is therefore also at the terminus.

CHAPTER 3

MACROMOLECULES

CHAPTER SUMMARY

All cells contain small molecules, whose molecular weights do not exceed a few hundred, and macromolecules, whose molecular weights range from 10^4 to 10^{12}. Macromolecules are of three major classes: proteins, nucleic acids, and polysaccharides. Proteins are polymers of 20 distinct amino acids; nucleic acids are polynucleotides, and polysaccharides are polymers of sugars. Amino acids can be thought of as a carbon atom (the α carbon) to which are attached three groups: an amino group, a carboxyl group, and a side chain. In a polypeptide chain amino acids are linked via a bond between an amino group of one amino acid and a carboxyl group of the adjacent amino acid; the resulting bond is called a peptide bond. At one end of a polypeptide chain is a free amino group; at the other end is a carboxyl group. The monomer of which a polynucleotide is built is a nucleotide, which consists of an organic base (either a purine or pyrimidine) linked to a five-carbon sugar (either ribose or deoxyribose), which is in turn linked to a phosphate group. The purines used are adenine (A) or guanine (G), and the pyrimidines are thymine (T), cytosine (C), or uracil (U). In DNA the sugar is deoxyribose and the pyrimidines are T and C; in RNA the sugar is ribose and the pyrimidines are U and C. In a linear polynucleotide chain the phosphate joins the 3' carbon of one sugar with the 5' carbon of the adjacent sugar, forming a phosphodiester group.

In aqueous solution single-stranded polymers are not extended but are folded to a great extent. In the absence of internal interactions a polymer will assume the random coil conformation. However, numerous interactions occur in most polymers: for example, hydrogen bonds, hydrophobic interactions, ionic bonds, and van der Waals interactions. Hydrogen bonds are weak noncovalent bonds between weakly charged groups. They are easily disrupted by heating and are not affected by the presence of various ions. Hydrophobic interactions occur between groups that interact poorly with water and thereby cause a clustering of such groups. Clustering of hydrocarbon and nonpolar-ring side chains of amino acids frequently occurs in proteins. A hydrophobic interaction is responsible for

the tendency of purines and pyrimidines to stack in nucleic acids. Hydrophobic bonds are strengthened by the addition of salts to the solution and are weakened by a variety of organic liquids. Ionic bonds occur between positively and negatively charged groups; such bonds are usually destroyed if the salt concentration is high. Van der Waals interactions are attractive forces between molecules or regions of molecules having complementary three-dimensional surfaces. An individual van der Waals bond is usually quite weak, but clusters of such bonds can be fairly strong. Each of the interactions just described tend to bring distant parts of polypeptide chains together.

Numerous methods are used to study macromolecules. Among the most common techniques are those that enable one to determine molecular weight and shape. An important technique is sedimentation induced by centrifugation. When centrifuged in solution, most molecules move at a rate determined by their shape and molecular weight. The ratio of the sedimentation velocity to the centrifugal force, which is known as the sedimentation coefficient and abbreviated *s*, is a characteristic of a molecule in a solution of a particular composition. The most common type of sedimentation is zonal centrifugation, in which a small sample of a solution of a molecule is layered on the top of a concentration gradient and the molecules sediment through the gradient. Fractionation of the centrifuge tube by drop collection enables one to determine the position of the molecules and to isolate them. Equilibrium centrifugation in a density gradient allows the separation of molecules based on their density rather than size and shape; this technique is important in the study of the nucleic acids. Electrophoresis is a valuable technique for separating molecules and is one of the most prevalent techniques in modern molecular biology. In electrophoresis molecules are placed in an electric field and move at a rate determined by their electric charge and their size. The most common type of electrophoresis is gel electrophoresis in which molecules are caused to move through a gel by an electric field. Gel electrophoresis is typically carried out under conditions in which all molecules to be separated have the same ratio of charge to mass, in which case the rate of movement usually depends only on the molecular weight of the molecules. Both DNA and proteins are commonly separated by gel electrophoresis. By measurement of relative movement the molecular weights of both proteins and DNA molecules can be determined fairly accurately. DNA molecules consisting of fewer than 200 nucleotides and differing by only one nucleotide pair can be separated. Electron microscopy is a technique for direct visualization of macromolecules. It is used most commonly in determining the size and shape of nucleic acid molecules and in the study of subcellular particles. In the Kleinschmidt technique DNA molecules suspended in a solution of cytochrome c are spread on the surface of a salt solution. The molecules become partially extended on the surface and are easily seen after being coated with a heavy metal. In the negative-contrast technique molecules or particles are suspended in a solution containing a heavy metal ion and then placed on a solid support film. The solution is opaque to electrons but is partially transparent in the region of the molecule or particle. The dimensions of the object and some details of its surface can be visualized by this technique.

BOLD TERMS

amino terminus, backbone, base-stacking, carboxyl terminus, equilibrium centrifugation in a density gradient, gel electrophoresis, hydrophobic interaction, Kleinschmidt spreading technique, macromolecules, negative contrast technique, nucleoside, nucleotide, peptide bond, phosphodiester group, polypeptide, polypeptide chain, protein, random coil, SDS-PAGE technique, sedimentation coefficient, shadowing, staining, sucrose gradient centrifugation, svedberg, van der Waals radius, zonal centrifugation

ADDITIONAL INFORMATION

In the text, measurement of the molecular weight of large DNA molecules (more than 10^8) was not discussed. Several techniques are available for this size range. For very large molecules (or for midsize molecules when only an estimate of *M* is needed) the value of *M* can be determined by measuring the value of *s*. However, *s* is dependent on the DNA concentration for large molecules. To obtain a standard *s* value that can be related directly to *M*, *s* is measured at various concentrations and a value called s_0 is obtained by extrapolation of the measured values of *s* to zero concentration. Then the equation $s_0 = 2.8 + 0.00834M^{0.497}$, which applies to double-stranded DNA in 1 M NaCl, can be used to evaluate *M*. Other equations exist for single-stranded DNA, for RNA, and for other ionic conditions. A simpler method is to use a radioactive label for the DNA molecules of interest (so they can be detected at a very low concentration) and to mix these DNA molecules with other radioactive DNA molecules whose molecular weights are accurately known, and to sediment the mixture through a sucrose gradient. The logarithm of the distance *d* sedimented is proportional to log *M* for a variety of ionic conditions; thus, a plot of log *d* versus log *M* yields the value of *M*. An interesting method, the viscoelasticity procedure, is available for linear DNA molecules with $M > 10^9$. In this method a solution of DNA is stirred in a special vessel, and the molecules are stretched by the motion. The stirring is stopped and the stretched molecules return to their normal state, causing the stirrer to rotate backwards. The rate of this reverse motion can be related to the molecular weight of the largest DNA molecules in the sample, which is useful because large molecules are frequently broken by hydrodyamic shear, as explained in Chapter 4. The viscoelasticity method has been used to measure the size of the DNA molecule of eukaryotic chromosomes. The largest molecule measured in this way is from a chromosome of *Drosophila americana*, whose molecular weight is 8×10^{10}.

Molecular weights of polymers are frequently expressed in the scientific literature in a unit called the dalton, which equals the molecular weight of a hydrogen atom. This unit, which has never been adopted by international commissions on nomenclature, is not used in *Molecular Biology*, since by international agreement molecular weights are unitless.

DRILL QUESTIONS

1 What are the monomers of which proteins, nucleic acids, and polysaccharides are composed?

2 What are glycoproteins and lipoproteins?

3 What groups in an amino acid become linked in forming a polypeptide chain? What is the linkage called?

4 What is meant by a side chain of an amino acid?

5 Which amino acid has no side chain, that is, only a hydrogen atom as a side chain?

6 What is the difference between a polypeptide chain and a protein?

7 What chemical groups are present at the termini of polypeptide chains?

8 Do polypeptide chains contain amino and carboxyl groups at positions other than the termini?

9 Which amino acid is unable to form a proper peptide bond?

10 Only one amino acid side chain engages in covalent bonds between different polypeptide segments. Which one is it?

11 Which amino acids can engage in hydrogen bonding?

12 Which compound, a nucleoside or a nucleotide, contains a phosphate group?

13 Which bases in nucleic acids are purines and which are pyrimidines?

14 Which nucleic acid base is unique to DNA? to RNA?

15 How do ribose and 2'-deoxyribose differ?

16 What is the name of the bond formed between a N atom in a purine or pyrimidine and a C atom in ribose or deoxyribose?

17 In a nucleic acid which carbon atoms are connected by a phosphodiester group?

18 What chemical groups are at the end of a polynucleotide?

19 What symbol is used to distinguish the single-letter abbreviations of the nucleosides containing either ribose or deoxyribose?

20 Why is the structure of a random coil not constant in time?

21 In simple terms what is the nature of a hydrophobic interaction?

22 What atoms participate in hydrogen bonds in proteins and nucleic acids?

23 What term describes the tendency of a macromolecule to fold so nonpolar groups cluster?

24 Name the kinds of bonds in proteins in which each of the following amino acids might participate: cysteine, lysine, isoleucine, glutamic acid?

25 Base-stacking occurs in both single- and double-stranded nucleic acids. This is an example of what kind of interaction?

26 Why is a van der Waals interaction between only two atoms unstable?

27 Why is the strength of the van der Waals interaction between several atoms highly dependent on complementary shapes?

28 Why are ionic interactions destroyed when the salt concentration in the surrounding medium is high?

29 What term would describe a macromolecule whose monomers do not engage in either attractive or repulsive interactions and which has free rotation between the monomers?

30 Define the terms shadowing, staining, and negative contrast, as used in the electron microscopy of macromolecules.

31 What is the definition of the sedimentation coefficient and what units are used for its measurement?

32 Which would have a higher sedimentation coefficient, a rigid rod or a flexible rod, both having the same radius and same mass?

33 In principle, how do zonal centrifugation in a concentration gradient and equilibrium centrifugation in a density gradient differ?

34 All DNA molecules move in the same direction when electrophoresed. Is this generally true of proteins also? Explain.

35 In gel electrophoresis what causes the movement of DNA molecules to depend on *M*? Do larger molecules move more slowly or faster than smaller molecules?

36 In SDS-PAGE electrophoresis of proteins what two features of the technique cause protein molecules to move in a single direction?

37 What are the methods of choice for determining the molecular weights of very small and midsize DNA molecules?

ANSWERS TO DRILL QUESTIONS

1 Proteins, amino acids; nucleic acids, nucleotides; polysaccharides, sugars.

2 Glycoproteins and lipoproteins are proteins to which sugars and lipids, respectively, are covalently linked.

3 An amino group and a carboxyl group are linked to form a peptide group.

4 The group, other than the amino or carboxyl group that is covalently linked to the α carbon.

5 Glycine.

6 All polypeptide chains are proteins, though the term protein is often reserved for the longer chains. However, proteins may contain several polypeptide chains bound by noncovalent interactions.

7 An amino group at one end and a carboxyl group at the other.

8 Yes, because some of the side chains contain these groups. For example, both aspartic acid and glutamic acid have a free carboxyl group, and both asparagine and glutamine have a free amino group.

9 Proline.

10 The side chain of cysteine, which has a terminal SH group.

11 All, using the CO and NH groups in the peptide bond. In addition, all polar amino acids can form hydrogen bonds via their side chains.

12 A nucleotide is a nucleoside phosphate.

13 Purines: adenine, guanine. Pyrimidines: cytosine, thymine, uracil.

14 Thymine, DNA; uracil, RNA.

15 Ribose has an OH group on the 2' carbon, and 2'-deoxyribose has a H atom at the same position.

16 An *N*-glycosidic bond.

17 The 3' and 5' carbons.

18 One 3'-OH and one 5'-P group.

19 A lower-case d or r precedes the letter in the deoxy and ribo molecules; for example, dC and rC for deoxycytidine and cytidine, respectively.

20 A random coil has no intramolecular interactions. Collisions with solvent molecules causes continual fluctuations in the shape of the molecule.

21 Two molecules that are poorly soluble, that is, interact poorly with water, can reduce contact with water by clustering, because clustering minimizes the ratio of surface to volume. More precisely, clustering reduces the number of molecules that become ordered by the presence of a poorly soluble molecule

22 Carbonyl oxygen, hydrogen bound to nitrogen, ring nitrogen, hydrogen in an OH group.

23 A hydrophobic interaction.

24 Cysteine—disulfide bonds; lysine—ionic bonds and hydrogen bonds; isoleucine—hydrophobic bonds; glutamic acid—ionic bonds and hydrogen bonds. All could of course participate in van der Waals bonds.

25 A hydrophobic interaction.

26 The interaction energy is very near that of the energy of thermal motion at room temperature, so the two atoms are very easily separated from one another.

27 The interaction energy decreases with the sixth power of the distance between the atoms. Thus, an attractive force will exist only when the atoms are very close to one another. Noncomplementary shapes prevent the required close approach.

28 The salt ions cluster around the charged regions of the macromolecule and thereby shelter them from one another.

29 A random coil.

30 Shadowing and staining refer respectively to physical and chemical deposition of heavy atoms onto a macromolecule. In the negative contrast procedure transparent macromolecules are observed within a droplet of a solution of small molecules that absorb electrons strongly.

31 It is the ratio of the velocity of the sedimenting molecule to the centrifugal force. The unit is the svedberg or S.

32 A flexible rod encounters less friction when moving through a fluid and thus has a higher sedimentation coefficient.

33 In zonal centrifugation molecules move continuously until they reach the bottom of the centrifuge tube. In equilibrium centrifugation the density of the suspending liquid is adjusted to a value such that molecules come to rest in the centrifuge tube at a point at which their density equals the density of the fluid in the tube.

34 No, because proteins may have a net positive or negative charge.

35 The charge per nucleotide is the same for molecules of all sizes, so the net force on all molecules is in a single direction. The pore size of the gel determines the rate at which a molecule of a particular size can move through the gel. Larger molecules cannot penetrate the pores as easily as smaller molecules and hence move more slowly through the gel.

36 An SDS molecule, which has a negative charge is bound to each peptide bond. Since the number of charged amino acid side chains in a typical protein is quite small, the total charge of the molecule, when SDS is bound, is determined almost exclusively by the SDS and is negative.

37 Small molecules, electrophoresis. Midsize molecules, electron microscopy.

PROBLEMS

1 Why does isotopic substitution (for example, ^{15}N for ^{14}N) cause a change in the density of a molecule containing the atom?

2 Which direction does a DNA molecule move in electrophoresis, + to – or the reverse, and why?

3 Consider the chemical structure of the heptapeptide N-Ala-Trp-Ser-Pro-Leu-Ile-Gly-COOH. How many peptide bonds are there?

4 If a particular macromolecule that is very compact in 0.01 M NaCl expands considerably in 0.5 M NaCl, what forces are probably important in determining the overall size and shape of the molecule?

5 Amino acids can be separated by electrophoresis on paper. A pH is chosen to make some amino acids positively charged and others negatively charged. Thus, some amino acids will move toward the anode and others toward the cathode, their charge principally determining the direction and rate. Often, amino acids

having the same charge (such as alanine and valine) also separate; in other words, they move at different rates, though in the same direction. What determines the difference in rate and will alanine or valine move faster?

6 A mixture of different proteins is subjected to electrophoresis in three polyacrylamide gels, each having a different pH value. In each gel five bands are seen. (a) Can one reasonably conclude that there are only five proteins in the mixture? Explain. (b) Would the conclusion be different if a mixture of DNA fragments was being studied?

7 A virus contains 256 proteins, 64 having a molecular weight of 1800 and 192 having a molecular weight of 26,000. If the virus is disrupted and analyzed by SDS-gel electrophoresis, what would be the relative distances migrated and the relative areas of the bands?

8 In the Kleinschmidt technique for visualizing DNA molecules, explain why it is necessary to use a very low angle when shadowing the DNA with metal. Is the metal being deposited directly onto the DNA molecule? Would it suffice to deposit metal from above?

9 A reagent commonly used in negative contrast electron microscopy is phosphotungstic acid. In this molecule what is actually absorbing the electrons? How would the appearance of an empty phage head compare to that of a head filled with DNA?

10 Which particle would be more easily visible by negative-contrast electron microscopy, a particle that the heavy metal ion could not penetrate or one that was freely permeable to the ion? Assume both particles have the same size.

11 Support films for electron microscopy are always very thin layers of either plastic or pure carbon. These films are very fragile and frequently break. Stronger films could be made of metals such as chromium. Why would such metal films not be useful?

12 The density of DNA in CsCl is approximately 1.7 g/cm^3 and that of most proteins is approximately 1.3 g/cm^3. What would you expect to be the density of a typical bacteriophage which is 50% protein and 50% DNA?

ANSWERS TO PROBLEMS

1 The volume of an atom is determined by the size of the electron cloud, which is not influenced by the mass of the nucleus. Thus, the atomic mass increases without a change in volume and the density of the "heavier" isotope is higher.

2 The charge of DNA is determined primarily by the negatively charged phosphates. Thus, DNA moves from – to +.

3 There are only five peptide bonds because the C–N bond between serine and proline is not a peptide bond.

4 The compactness probably results from an ionic bond(s) between unlike charged groups.

5 Electrophoretic mobility increases with charge but is reduced by friction between the moving molecule and the solvent molecules. Valine has a longer side chain than alanine, so it encounters greater friction and moves more slowly.

6 (a) No, since differences in shape can affect mobility—that is, two proteins differing in both molecular weight and shape might have the same electrophoretic mobility. (b) Yes, for all fragments have the same basic shape and same charge-to-mass ratio.

7 The relative distances migrated are roughly the ratio of log 1800/log 2600, with the smaller molecule moving faster. The ratio of the areas is (192 x 26,000)/(64 x 1800) = 43.3 with the slower band having the larger area.

8 The DNA is first coated with molecules of cytochrome c, which increases the effective thickness of the DNA. Shadowing at a low angle allows large amounts of metal to accumulate against the thickened molecule in the manner that snow accumulates against a fence in a storm, forming a drift or shadow. Depositing metal from directly above would cover the molecule and the background uniformly and eliminate all potential contrast.

9 In the negative contrast technique, the biological sample (which is transparent to electrons) is surrounded by an opaque material—in this case, tungsten atoms. The intensity of the image is inversely proportional to the thickness of the opaque material. Hence, a phage head filled with DNA appears as a solid white object against a darker background, but an empty phage head will become filled with phosphotungstic acid and will therefore appear as a light hexagonal outline against a dark background.

10 The impermeable particle. The visibility of a particle is determined by its contrast with the surrounding material. When a particle is embedded in a medium opaque to electrons, the contrast will be determined primarily by the thickness of region that excludes the heavy metal ion.

11 Such metals would adsorb more electrons than the sample, so contrast would be diminished, if not lost.

12 (1.7 + 1.3)/2 = 1.5 g/cm^3.

MORE CHALLENGING PROBLEMS

1 Many proteins bind metal ions quite tightly. Which amino acid would probably be responsible for this binding? If the ion were Hg^{2+}, which amino acids would most likely be involved?

2 If a macromolecule that is a near-random coil and that is in 0.2 M NaCl becomes very extended and rigid in 0.01 M NaCl, what forces are probably involved in the extension?

3 Gel electrophoresis is done in solutions having low salt concentration. Why do you think this is necessary?

4 Suppose that you have isolated a protein that seems to have two enzymatic activities. This observation makes you suspect that you may have two proteins that purify together. To check this, you subject the preparation to electrophoresis in a polyacrylamide gel at a variety of pH values. In each gel, a single band results, but the band is sufficiently broad that it may contain two proteins which do not separate. In an SDS gel, a single, broad band is also found. Because a protein with two enzymatic activities is rare, it is necessary to try a little harder to see if the breadth of the band is due to the presence of two components. What parameters could you vary to improve resolution by electrophoresis? What other methods (nonelectrophoretic) might you try?

5 An enzyme has a sedimentation coefficient of 16S at pH 7. At pH 3 the sedimentation coefficient drops to 11S and at the same time the enzyme loses all activity (even when the pH is restored to 7). It is known that there is no change in molecular weight at pH 3. (a) What is the approximate shape of the protein at pH 7? (b) If the first value was 4S instead and the value at pH 3 was 10S, what would you conclude?

ANSWERS TO MORE CHALLENGING PROBLEMS

1 Those with charged groups other than the α-amino carboxyl group bind metal ions. An ionizable protein can bind a metal ion simply by forming an ionic bond. Many metal ions (such as Mg^{2+} and Cu^{2+}) form complexes with amino groups. The Hg^{2+} ion binds SH groups and is frequently bound to methionine and cysteine.

2 The molecule probably contains many groups having like charge, and in the lower ionic strength of 0.01 M NaCl these groups repel one another.

3 If the salt concentration is high, the electric current is high and the temperature will rise until the gel melts.

4 Vary the pore size of the gel so that the dependence of the mobility on the molecular weight would increase. For those of you who are familiar with chromatographic procedures, gel chromatography, which is sensitive only to molecular size, and ion-exchange chromatography, which separates proteins primarily by their charge, would also be useful.

5 (a) The molecule is compact and probably is roughly spherical. (b) The molecule is an extended rod.

CHAPTER 4

NUCLEIC ACIDS

CHAPTER SUMMARY

Except for certain viruses, the genetic material of all organisms is DNA. In prokaryotes the DNA is dispersed throughout the cell. In eukaryotes the major fraction is enclosed in the nuclear membrane; some DNA is also found in cytoplasmic organelles, such as mitochondria and chloroplasts, when they are present. DNA was shown to be the genetic material, in contrast with proteins, through (1) the genotypic and phenotypic transformation of bacteria by isolated DNA and (2) the injection of phage DNA, but not phage protein, into bacteria that subsequently produce progeny phage.

DNA is a double-stranded polymer consisting of repeating nucleotides. A nucleotide has three components—a base, a sugar (which is deoxyribose in DNA), and a phosphate. Sugars and phosphates alternate in forming a single polynucleotide chain that has distinct termini, a 3'-hydroxyl group and a 5'-phosphate group. In double-stranded DNA the two strands are antiparallel: each end of the double helix carries one3'-OH group and one 5'-P group. Four bases are found in DNA: adenine and guanine (purines) and cytosine and thymine (pyrimidines). Equal numbers of purines and pyrimidines are found in double-stranded DNA. The bases are paired; only pairs between adenine and thymine (an AT pair) and guanine and cytosine (a GC pair) are present. This pairing is responsible for holding the two polynucleotide strands together in a double helix. The base composition of DNA varies from one organism to the next.

The molecular weight of DNA molecules varies enormously, from small plasmids (about 10^6) to human chromosomes (nearly 10^{11}). Molecules above a molecular weight of about 50×10^6 are highly susceptible to breakage by hydrodynamic shear forces encountered in ordinary laboratory operations.

The double helix can assume a variety of forms. That which is common in solution and most likely the intracellular form is the B form, in which the base pairs are parallel to one another, separated by 3.4 angstrom units, and in which there are roughly 10 base pairs per helical turn. This structure has two external

grooves, the major groove and the minor groove. The major groove is the site of protein binding, as will be seen in Chapter 7. A left-handed helical form of DNA, Z-DNA, also exists. This form is found only in regions in which purines and pyrimidines alternate along a single strand and under certain ionic conditions. Whether it exists intracellularly is widely debated.

Hydrophobic interactions cause DNA base pairs to stack one above another. In the B form this stacking maintains the bases in a parallel array in which each pair is perpendicular to the helical axis. This array puts all bases in a position in which hydrogen-bonding is possible. Treatment of DNA with any agent or reagent that either disrupts the orientation of the bases or reduces the ability to hydrogen-bond causes separation of the two polynucleotides and collapse of each strand. Such separation is called denaturation. Common denaturating agents are high temperature and pH > 11.3. Useful denaturating reagents are formaldehyde and formamide, both of which interfere with base-pairing, and various alcohols, ureas, and other organic compounds, which break down hydrophobic interactions. Denaturation is frequently studied by measuring the optical absorbance of a DNA solution as a function of temperature, pH, or concentration of some reagent. The basis for this measurement is the lower absorbance of double-stranded DNA compared to single-stranded DNA; that is, DNA is hypochromic. A curve that relates absorbance to the denaturing agent is called a melting curve. The midpoint of a thermal melting curve is designated T_m; the value of T_m increases with GC content and decreases with salt concentration. Individual hydrogen bonds are not particularly strong and at room temperature are constantly being broken and re-forming, a phenomenon called breathing. Proteins that bind tightly to single-stranded DNA can enter the DNA structure during an act of breathing and remain there. Some of these proteins bind tightly to one another and hence are able to invade the DNA helix; these are called melting or helix-destabilizing proteins.

When denatured DNA is restored to conditions in which double-stranded DNA is the preferred form, single strands can re-form a double helix. This process, which is called renaturation, requires conditions that eliminate randomly formed base pairs (for example, a temperature near but still less than T_m) and that allow the highly charged DNA strands to approach one another (high salt concentration). Renaturation is a concentration-dependent process, and the rate of renaturation can be used to determine the relative concentrations of different base sequences in a heterogeneous DNA sample. Cot analysis enables one to determine the number of copies of particular sequences and the number of bases in each sequence. Cot analysis has demonstrated that most subsequences of the single DNA molecule of a prokaryote occur only once. In contrast, subsequences of the DNA of eukaryotes fall into several classes, based on the degree of repetition; these classes are the unique (1 copy per genome), slightly repetitive (1 to 10 copies), middle repetitive (10 to hundreds), and highly repetitive (up to several million copies) sequences. Satellite DNA is a special class of highly repetitive DNA in which a short sequence has a density in CsCl solutions that is quite different from that of the bulk of the DNA of the organism. One class of satellite DNA is a short sequence present in heterochromatic regions of eukaryotic chromosomes; these are regions that are condensed and visible throughout the life cycle of the cell, that is, even when chromosomes are not normally visible. These heterochromatic regions are often located in the centromere (site of attachment of chromosomes to the mitotic spindle) and telomeres (termini of chromosomes).

In prokaryotes and in many eukaryotic viruses DNA molecules are usually circular and supercoiled. These twisted molecules are all underwound and are called negative superhelices. In prokaryotes supercoiling is primarily a result of the activity of topoisomerases, for example, *E. coli* DNA gyrase. In eukaryotes supercoiling results from binding of DNA-binding proteins, as will be described in Chapter 7. Because of the underwinding, supercoils usually contain single-stranded regions that are larger and have a greater lifetime than those produced in linear molecules by breathing. Supercoiled molecules are under great torsional strain and introduction of one single-strand break causes relaxation and complete loss of superhelicity. Supercoiling can be detected by observing (1) the density shift in CsCl solutions containing the intercalating agent ethidium bromide or (2) the mobility during agarose gel electrophoresis compared to linear and nicked circular DNA.

DNA contains a variety of sequences which are duplicated and adjacent. An inverted repeat has the sequence along one strand of ABCDEE'D'C'B'A', in which a prime denotes base-pairing complementarity. Such a sequence is also called a palindrome. A direct repeat has the form ABCDEABCDE. Inverted repeats often contain a spacer and have the form ABCDE-WXYZ-E'D'C'B'A'. A single strand having an inverted repeat can fold back on itself and form a hairpin and, if a spacer is present, a stem and loop. It is thought that a double-stranded DNA molecule can form a cruciform in the region of an inverted repeat.

RNA is usually a single-stranded molecule, although often as many as half of the bases are internally paired. Thus, a typical RNA molecule consists of double-stranded segments separated by single-stranded loops and has a well-defined secondary structure. The base-paired regions rarely contain more than 50 base pairs.

A variety of reagents and enzymes can break bonds in nucleic acids. For example, *N*-glycosidic and phosphodiester bonds are broken at pH 1 and 4, respectively. The phosphodiester bond of RNA, but not of DNA, is broken at pH > 11. Nucleases are enzymes that break phosphodiester bonds. Some are non-specific but many are active only on DNA or on RNA. Some act only on the terminal internucleotide linkage (exonucleases); others act on any phosphodiester bond (endonucleases). Nucleases may also recognize single- or double-stranded DNA specifically.

BOLD TERMS

apurinic acid, base composition, breathing, centromere, complexity, covalently closed circle, cruciform, denaturation, denatured, direct repeats, DNA gyrase, DNA-RNA hybridization, dyad symmetry, endonuclease, ethidium bromide, euchromatin, exonuclease, fluorescent antibody, genome, hairpin, helix-destabilizing protein, heterochromatin, heteroduplexing, highly repetitive, hybridization, hyperchromic, hypochromic, *in situ* hybridization, inverted repeat, inverted repetition, linking number, major groove, melting curve, melting protein, melting temperature, membrane filter, middle repetitive, minor groove, native, nicked circle, nicks, node, nuclease, palindrome, reannealing, renaturation, renatured DNA, repeated base sequences, satellite DNA, slightly repetitive, S1 endonuclease, stem-and-loop, supercoil, superhelix, superhelix density, T_m, telomeres, twisting number, unique, writhe, Z-DNA

ADDITIONAL INFORMATION

In the text it was pointed out that all naturally occurring superhelices are left-handed. While *Molecular Biology* was in production, a report appeared in the current literature of a positively supercoiled molecule present in a viruslike particle contained in the bacterium *Sulfolobus acidocaldarius* . It is interesting that such an unusual molecule would be found in this bacterium since *Sulfolobus* is not a typical prokaryotic bacterium but belongs to a separate group called the archaebacteria. An enzyme, reverse gyrase, that produces positive supercoiling was isolated from the same bacterium. It can convert both nonsupercoiled and negatively supercoiled DNA into positive supercoils. This report can be found in *Nature, 321*, 256, 1986.

Most physical studies indicate that DNA is a fairly rigid molecule in which severe bending would require breakage of hydrogen bonds. Nonetheless the molecule must be bent considerably in many situations—for example, when a molecule 1 mm long is folded into a bacterium of length 0.002 mm, or when packaged in a phage head. Recent experiments show that bending of DNA molecules can occur in regions including certain base-sequence patterns. The sequence

AATTCCC*AAAAA*TGTC*AAAAA*TAGGC*AAAAAA*TGCC*AAAAA*TCCAA

in which there are recurring runs of A, is bent. Studies of related sequences has shown that the length of the block of A's is important, 6 A's being optimal and 3 A's insufficient. Furthermore, the blocks of A's must be in phase with the helix in the sense that the length of the A tract plus the length of the intervening sequence must be 10 bases (one turn of the helix). Bending occurs at both ends of the tract of A's. This report can be found in two articles: *Nature, 320*, 487 and 501, 1986.

DRILL QUESTIONS

1 Name the bases in DNA and in RNA.

2 What base pairs with G? with A?

3 Which DNA bases are pyrimidines and which are purines?

4 Which sugars are in DNA and in RNA, and by what chemical feature do they differ?

5 How many phosphate groups are there per base in DNA and in RNA?

6 What is the base sequence of the strand complementary to 5'-GCAATCCG-3'?

7 Of the molecules adenine, adenosine, adenylic acid, which is a base, a nucleoside, and a nucleotide?

8 With what base will the modified base 5-methylcytosine pair?

9 How many hydrogen bonds are there in an AT and a GC base pair?

10 What chemical groups are at the end of a single polynucleotide strand?

11 How many turns of a double helix are there in a molecule consisting of 45 base pairs?

12 How many bases are in a molecule whose length is 2 micrometers?

13 Are most naturally occurring superhelices underwound or overwound?

14 Using the standard convention for writing polynucleotides, which end of GAC-TAAC is the 5' end?

15 In DNA, what is the relation between the sums ([A] + [G]) and ([T] + [C]), in which [ ] denotes concentration?

16 When referring to helices, what is meant by right-handed and left-handed? What is the handedness of a normal DNA helix?

17 In what sense are the two strands of DNA antiparallel?

18 Distinguish the terms denaturation and renaturation with respect to DNA.

19 What term is used to describe a DNA sequence (in one of the complementary strands) such as AGACGTCT?

20 What is meant by a nuclease, and how do endonucleases and exonucleases differ?

21 What structure could be assumed in the region of a single-stranded DNA molecule in which the sequence of one strand is AGGTCTCCTAGACCT?

22 Can an exonuclease depolymerize a supercoiled DNA molecule?

23 What is meant by high-GC DNA?

24 Two DNA molecules have different densities when centrifuged to equilibrium in CsCl. What possible explanations may be given?

25 What is meant by the optical absorbance of a solution of a macromolecule? Is the absorbance per mole of nucleotides of single-stranded or double-stranded DNA higher?

26 What is meant by base stacking?

27 Which of the following operations will induce formation of a supercoil: joining the ends of a linear DNA and doing nothing else; twisting the two ends of a linear DNA and then joining the ends together; joining the ends of a linear DNA and then twisting the circle?

28 Why does a supercoil have a higher density in buoyant CsCl than a linear molecule, when ethidium bromide is present?

ANSWERS TO DRILL QUESTIONS

1 DNA—adenine, guanine, cytosine, thymine. RNA—adenine, guanine, cytosine, uracil.

2 G with C, and T with A.

3 Thymine and cytosine are pyrimidines, and adenine and guanine are purines.

4 DNA—deoxyribose. RNA—ribose. Ribose has an OH group on the 2' carbon and deoxyribose has an H.

5 One.

6 3'-CGTTAGGC-5'.

7 Adenine—base; adenosine—nucleoside; adenylic acid—nucleotide.

8 Guanine, since none of the hydrogen-bonding groups are methylated.

9 AT, 2; GC, 3.

10 A 5'-phosphate and a 3'-hydroxyl group.

11 45/10 = 4.5.

12 The linear density of DNA is roughly 2×10^6 molecular weight units per

micrometer, and the average weight of a base pair is 660. Thus, the number of base pairs is $4 \times 10^6/660 = 6060$. One could also calculate it from the fact that the base pairs are separated by 3.4 angstroms.

13 Underwound.

14 The terminus with the G.

15 The first is the sum of the molar concentration of the purines and the second is the sum for the pyrimidines; thus, the two sums are equal.

16 If a viewer watched an object move along a right-handed helix, the object would follow a clockwise motion and move away from the viewer; a counterclockwise motion would cause the object to move away in a left-handed helix. The DNA helix is right-handed.

17 The sugars have opposite orientations in the two strands; thus, at one end of a double-stranded DNA molecule one strand terminates with a 3'-OH group and the other strand with a 5'-P group.

18 Denaturation, loss of double-stranded structure and separation of the individual strands.

19 It is an inverted repeat or a palindrome.

20 A nuclease is an enzyme capable of breaking a phosphodiester bond. An endonuclease can break any phosphodiester bond, whereas an exonuclease can only remove a terminal nucleotide, or sometimes a terminal oligonucleotide.

21 A stem-and-loop in which the loop is TCCTA.

22 It could not because there would be no free ends, which an exonuclease requires. The supercoiling is irrelevant; only the fact that it is covalently closed is important.

23 A molecule in which the fraction of base pairs that are GC pairs is considerably greater than 1/2.

24 The simplest is that they have different G+C contents. Another possibility is that one has a molecule covalently linked to it whose density in CsCl differs from that of DNA; occasionally, sugars and proteins have been observed linked to DNA.

25 Absorbance is the negative logarithm of the ratio of the transmitted light to the incident light. Single-stranded DNA has a higher molar absorbance than double-stranded DNA.

26 Base stacking is the tendency for bases in DNA or RNA to be oriented such that their planes are parallel and such that one base is directly over or nearly over the next base in the chain.

27 Only twisting followed by joining.

28 Less ethidium bromide can be bound by a supercoil than by a linear molecule. Thus, the decrease in the density of a supercoil is less than that of a linear molecule.

PROBLEMS

1 A technique is used for determining base composition. Rather than giving mole fractions for individual bases, it yields the value of [A]/[C]. If this value is 1/3, what fraction of the bases are A?

2 Analysis of a DNA sample from a bacterium indicates that 18% of the bases are A. What fraction is C?

3 Analysis of the DNA extracted from a newly isolated virus indicates the molar composition: 40% A, 20% T, 15% G, 25% C. (a) What is the base composition of the DNA, based on this analysis? (b) How should this result be interpreted, in terms of the structure of the viral DNA?

4 *E. coli* DNA has a molecular weight of about 2.7×10^9. How long is this molecule and what does this tell you about the state of the intracellular DNA? (The length of a typical rodlike bacterium is 10^{-4} cm.)

5 Could a single-stranded DNA molecule with base sequence 5'-GATTGCC-GGCAATC-3' fold back on itself to form a hairpin?

6 What would be the approximate molecular weight of a DNA molecule whose length is 16.4 micrometers?

7 A typical gene utilizes a segment of DNA whose molecular weight is 10^6. How many turns of the helix does this represent?

8 An elegant combined chemical and enzymatic technique enables one to isolate "nearest neighbors" of bases. That is, the DNA is broken at random, and all possible adjacent nucleotides are isolated. For example, if the single-stranded tetranucleotide 5'-AGTC-3' were treated in this way, the nearest neighbors would be AG, GT, and TC (they are always written with the 5' terminus at the left). Before techniques were available for determining the base sequence of DNA, nearest-neighbor frequency analysis was used to determine sequence relationships. Nearest-neighbor analysis also indicates an important feature of DNA structure, namely, its antiparallel structure, which you are asked to examine in this problem by predicting some nearest-neighbor frequencies.

Assume that you have determined the frequencies of the following nearest neighbors: AG, 0.15; GT, 0.03; GA, 0.08; TT, 0.10. (a) What must the nearest-neighbor frequencies of CT, AC, TC, and AA be? (b) If DNA had a parallel, rather than an antiparallel, structure, what nearest-neighbor frequencies would you know and what would they be?

9 Order the DNA molecules shown below from lowest to highest melting temperature. In each case only *one* of the complementary strands is given.

1. AAGTTCTCTGAA 2. AGTGTCAATGCAG 3. GGACCTCTCAGG

10 Which of the following DNA molecules would have the lower temperature for strand separation? Why? In each case only *one* of the complementary strands is shown.

1. AGTTGCGACCATGATCTG 2. ATTGGCCCCGAATATCTG

11 The DNA molecules shown below are denatured and then renatured. In both cases, renaturation takes place at a single temperature. Which molecule will renature with greater probability to form the original structure? Why? (Hint: Think about intrastrand interactions.) In each case only *one* of the complementary strands is shown.

1. ATATATATAT 2. TAGCCGATGC

12 The following DNA molecules are denatured and then renatured. It is found that one requires a higher temperature for renaturation. Which one is it? In each case only *one* of the complementary strands is shown.

1. GAGCTGCATCAGATGCAG 2. ATCGGGGTACCCCGATAA

13 When a DNA solution is boiled, the two strands of each of the DNA molecules separate because the hydrogen bonds break. When the solution is cooled again, the two strands do not usually rejoin, because the probability is very low that complementary strands will collide in such a way that complementary base sequences match. In other words, boiled DNA remains single-stranded, although under certain conditions (which are not considered in this problem) the strands will reassociate. Consider a DNA molecule containing only adenine and thymine, in which the bases alternate along each strand, that is, the sequence is ATATATAT.... If such a molecule were boiled and then cooled, what types of base-paired structures might be formed by the individual strands?

14 ^{15}N-labeled DNA from phage T4 is mixed with T4 DNA of normal density. The solution is then heat-denatured and renatured. The resulting DNA is analyzed by centrifugation in a CsCl density gradient. How many bands will be observed and what will their relative proportions be?

15 If ^{15}N-labeled T4 DNA is mixed with ^{14}N-labeled DNA from phage P2 and then boiled and cooled, two bands are seen in a CsCl density gradient. Explain this number of bands.

16 A double-stranded hexanucleotide, for example, 3 guanines in one strand and 3 cytosines in the other, has very low thermal stability compared to a polynucleotide containing 1000 guanines in one strand and 1000 cytosines in the other. Explain the difference.

17 Suppose you had a method for measuring unpaired bases in DNA and obtained the data shown in the table below at a particular temperature. How many pairs of bases would be unpaired in a polynucleotide having 100 base pairs? Where would you expect these unpaired bases to be? Why will they be in that position? (Hint: see Problem 16.)

Number of base pairs in the DNA	*Fraction unpaired, (percent)*
5000	0.28
1000	1.40
250	5.60

18 When DNA is placed in distilled water, the two strands come apart. Explain. (Consider the effect of ions in solution on the interaction between the charged groups.)

19 Consider a protein that is very polar (that is, highly charged) but has a small, very nonpolar region. The nonpolar region binds very tightly to the bases in DNA. Which of the following results would follow addition of this protein: (1) the DNA would precipitate; (2) the two strands would come apart; (3) the DNA would become more stable.

20 DNA containing only G in one strand and only C in the other strand can be prepared enzymatically in the laboratory. It is quite difficult to denature these molecules because the melting temperature is higher than 100 C. However, with the addition of denaturants, strand separation can be accomplished. Assuming that you had a sample all of whose molecules contained the same (or nearly the same) number of bases, what would you expect to be the product when a denatured sample is renatured?

21 If a DNA solution with an absorbance at 260 nm of 1.00 is boiled, the absorbance increases to 1.37. This is an indication of the conversion of double-stranded to single-stranded DNA. When returned to 25 C, the absorbance remains the same, if the DNA is in a solution of low ionic strength and contains no polyvalent

cations (for example, 0.01 M NaCl), but drops to 1.12 if the ionic strength is high (for example, 0.5 M NaCl). (a) Explain the observation. (b) By various chemical means (e.g., by treatment with nitrous acid) the two bases in a base pair can be covalently joined. Predict the absorbance changes when this DNA is boiled in 0.5 M NaCl and then returned to 25 C. (c) Predict absorbance changes for a double-stranded deoxynucleotide containing G in one strand and C in the other. Would the change be the same for a circular polymer? Assume the ionic strength is high. (d) A double-stranded DNA molecule whose base sequence in one strand is ATGCATATATGCAT is boiled in 0.5 M NaCl and then returned to 25 C. What will be the ratio of the values of the absorbance before boiling and after a cycle of boiling and cooling?

22 Denaturants can be ranked in terms of the number of degrees of lowering the melting temperature by addition of the substance to yield a concentration of 0.1 M (even though the decrease might be only a few degrees). For each of the following pairs, which substance is probably the more effective denaturant and why:

(a) Ethanol, propanol.
(b) Urea, dimethylurea.
(c) Dimethyl ether, diethyl ether
(d) 1-amino-2-hydroxypropane, 1,3-diamino-2-dihydroxypropane.

23 At pH 12, strands of DNA separate because hydrogen bonds break. What is the cause of the breaking of hydrogen bonds? Why are the hydrophobic forces unable to maintain the double-stranded structure?

24 It has been stated that a single-stranded polynucleotide is helical if bases are stacked, and this stacking tendency has been used to explain the stability of DNA. Since stacking in a single-stranded polynucleotide is between adjacent bases, how can stacking explain the tendency of the two strands of the double helix to stay together? (The principle here is the same as that in Problem 23).

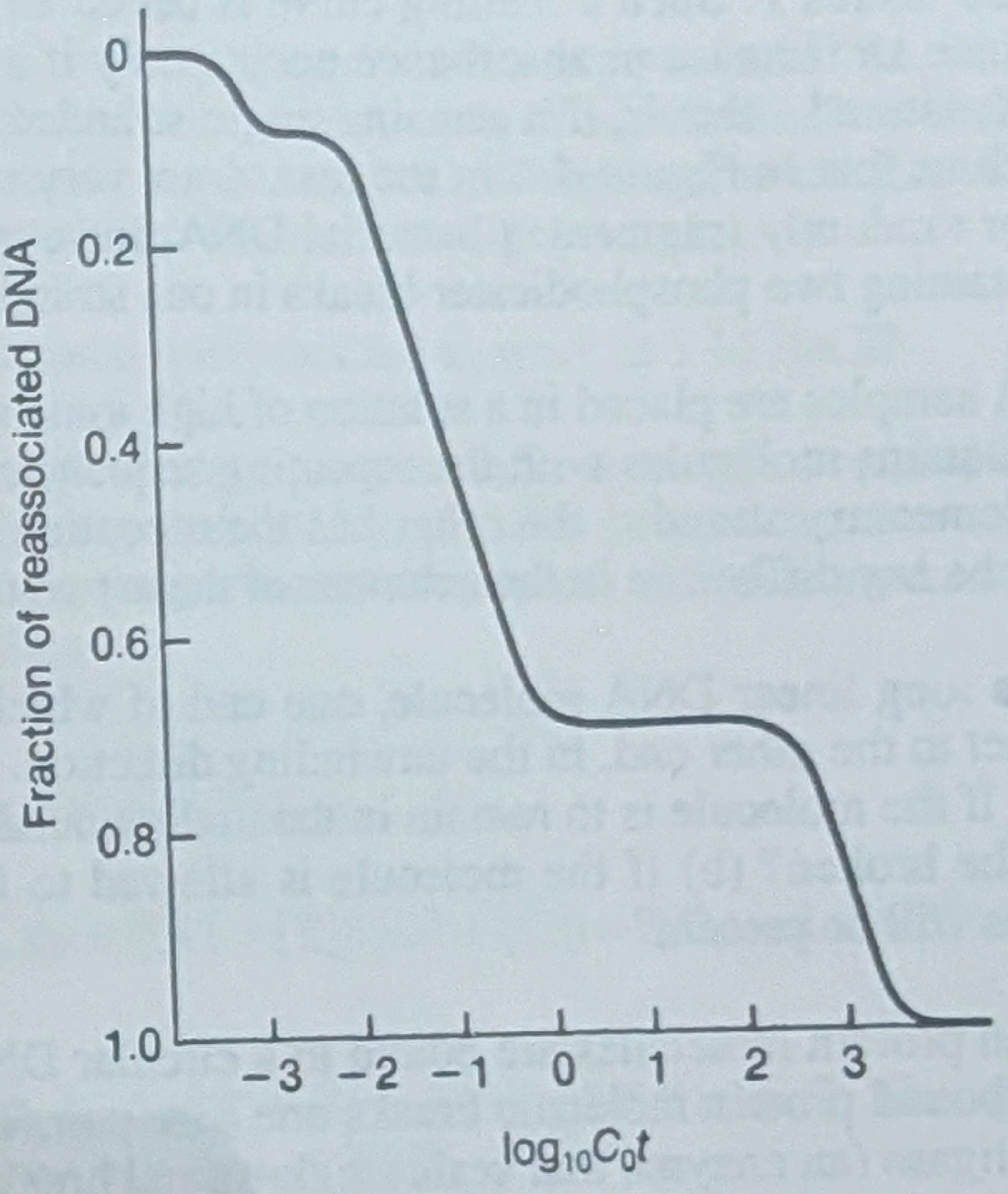
25 When double-stranded DNA is heated, the absorbance at 260 nm increases to a value that is 37% higher than the absorbance at 25 C. However, in a solution of 7 M sodium perchlorate, the maximum value is only 26% higher than that at 25 C. Explain.

26 Suppose you have obtained a melting curve having the normal shape, but the total increase in absorbance is only 20%. How might you explain the low value? Would the determination of the melting temperature from such a curve be believable?

27 You have just prepared two DNA samples, one native and one denatured, and have dialyzed each against 0.01 M NaCl. You do not know the values of the absorbance at 260 nm before dialysis. You then add a very small amount of an enzyme to each and in so doing mix up the samples. How can you determine the identity of each sample by an absorbance measurement? You may assume that a small part of each sample is consumed in the testing and that the enzyme will not interfere with the test.

28 Why does salt concentration affect the melting temperature of DNA?

29 A Cot curve is shown below. (a) What fraction of the DNA is contained in components that are unique? What fraction is middle repetitive? What fraction is highly repetitive? (b) If the unique sequences are represented once per genome, how many copies of each of the other two classes exist in the genome?



30 Bacterial DNA can be density-labeled if the bacteria are grown in a medium containing a heavy isotope (such as ^{15}N or ^{13}C). If both strands are so labeled, the DNA is said to be heavy (HH) as opposed to normally light (LL) DNA. If equal amount of HH and LL DNA are mixed, heated to 100 C, and slowly cooled to allow renaturation to occur, 25% of the DNA will be HH, 25% will be LL and 50% will be HL (hybrid). Suppose 45 micrograms of HH DNA and 5 micrograms of LL DNA are mixed, heated to 100 C, and renatured, how many micrograms of HH, HL, and LL will result?

31 DNA from organism A, labeled with ^{14}N, is hybridized with an equal concentration of DNA from organism J, labeled with ^{15}N, and then centrifuged to equilibrium in CsCl. If 5% of the total renatured DNA has a hybrid density, what fraction of the base sequences is common to the two organisms?

32 A linear phage DNA molecule is hybridized with a molecule that is identical except that the central 10% of the molecule is deleted. (a) What is the structure of the heteroduplex that would be seen by electron microscopy? (b) If the missing 10% is replaced by a piece of nonhomologous bacterial DNA having the same length, what will the structure of the heteroduplex be? How will this differ if the nonhomologous bacterial DNA is only half the length of the deleted phage DNA?

33 When melting curves for DNA solutions are obtained, the indicated temperature is generally that at which the absorbance was measured. Another kind of melting curve can be obtained by heating the DNA to a temperature *T*, cooling the DNA to 25 C, then measuring the absorbance at 25 C and plotting this value of the absorbance versus *T*. Such a melting curve is called an irreversibility curve or i-curve since an increase in absorbance occurs only if a DNA molecule is irreversibly denatured—that is, if it remains single-stranded. For a standard melting curve such as that in Figure 4-6 in the text, draw very carefully a hypothetical i-curve for randomly fragmented bacterial DNA, intact phage DNA, and phage DNA containing two phosphodiester breaks in one strand.

34 Two DNA samples are placed in a solution of high ionic strength, for example, 2 M. One contains molecules with the repeating sequence 5'-ATGC-3' (in one of the complementary strands; the other has the repeating sequence 5'-AGTC-3'. Will there be any difference in the behavior of these two molecules?

35 Consider a long linear DNA molecule, one end of which is rotated four times with respect to the other end, in the unwinding direction. The two ends are then joined. (a) If the molecule is to remain in the underwound state, how many base pairs will be broken? (b) If the molecule is allowed to form a supercoil, how many nodes will be present?

36 Suppose ten protein molecules are bound to a circular DNA molecule having a nick. Each bound protein molecule breaks one base pair. The nick is then sealed with DNA ligase (an enzyme that seals single-strand breaks having a free 3'-OH and 5'-P group) and the protein molecules are removed. What will be the shape of the DNA molecule after removal of the protein?

37 A covalently closed circular DNA molecule that is relaxed (neither positively nor negatively supercoiled) is partially denatured and 50 base pairs are broken. How does the degree of supercoiling change?

38 Could an endonuclease specific for single-stranded DNA attack a supercoiled molecule?

39 DNA molecules are easily broken by hydrodynamic shear forces. These forces act in such a way that a DNA molecule breaks near its middle. By slowly increasing the magnitude of the force, a molecule can be successively broken into halves, then quarters, then eighths, and so on. Let us call the values of a force just sufficient to break a linear molecule of mass *m* into halves F(1/2); for quarters and eighths we use the higher values F(1/4), F(1/8); and so on. If you have a circular molecule of mass *m*, approximately what force is needed to break the molecule? What is the size of the fragments?

b) 40 Would you expect the density decrease in CsCl induced by the presence of ethidium bromide to be affected by the G+C content?

41 In CsCl containing ethidium bromide, a supercoiled and a linear DNA have densities 1.592 and 1.556 g/cm^3, respectively. What would be the density of a supercoil linked (as in a chain) with an open circle of (a) identical mass or (b) twice the mass? (Linked circles do exist and are called catenanes).

42 The sedimentation velocity properties of a supercoiled DNA molecule are being studied as a function of the concentration of added ethidium bromide. It is found that *s* decreases, reaches a minimum, and then increases. Explain. (Hint: Recall that *s* is a function of both molecular weight and shape.)

43 Why do you think most nucleases fail to work in 1 M NaCl?

44 The enzyme S1 nuclease acts preferentially on single-stranded DNA. The activity against double-stranded DNA molecules is weak but increases with temperature even when the temperature is much less than that producing an increase in absorbance. Explain.

ANSWERS TO PROBLEMS

1 If [A]/[C] = 1/3, then ([A] + [T])/[G] + [C]) = 1/3, and the DNA is 25% A+T, or 12.5% A.

2 If 18% is A, 18% is also T, for a total of 36% for A+T. Therefore, G+C will be 64%, and C will be half that, or 32%.

3 (a) The base composition is not meaningful in terms of the standard definition because of the lack of complementarity. Therefore, the individual percentages would normally be stated. (b) Clearly, [A] is not equal to [T], nor is [G] equal to [C]. Thus, the DNA cannot be double-stranded and is probably single-stranded. Possibly, it has both double-stranded and single-stranded regions.

4 From the molecular weight of a nucleotide pairs (660) the number of pairs is about 4×10^6. The spacing between nucleotide pairs is 3.4 angstroms = 3.4×10^{-8} cm, so the length is $(4 \times 10^6)(3.4 \times 10^{-8}) = 0.136$ cm. The DNA is circular so if it were stretched out, it would be (1/2)1360 = 680 times the length of the bacterium. Thus, the DNA must be highly folded in the cell.

5 Yes. The base-pairing rule would be satisfied perfectly and the two components

of the double-stranded portion will be antiparallel.

6 The mass/length of double-stranded DNA is approximately 2×10^6 per micrometer; hence, the molecular weight is 32.8×10^6.

7 Each turn of the helix has about 10 base pairs or a molecular weight of 6700. Thus, there are $10^6 / 6700 = 149$ turns.

8 (a) To answer this question we must note the convention used in stating polynucleotide sequences: the 5' end is written at the left. Thus, the complementary sequence to AGTC (which is really 5'-AGTC-3') is GACT (5'-GACT-3'). Hence, the frequency of CT is the same as that of its complement AG, which is given as 0.15. Similarly, AC = 0.03, TC = 0.08, and AA = 0.10. (b) If DNA had a parallel structure, both ends of each strand would be the same (say, 5'), so the sequence complementary to AGTC would be TCAG. Thus, if AG = 0.15, TC will also be 0.15. Similarly, CA = 0.03, CT = 0.08, and AA = 0.10.

9 The order should be 1, 2, 3, since melting temperature increases with increasing (G+C) content.

10 Molecule 1 because of the long tract of GC pairs in 2.

11 Molecule 2, because the strands in 1 will preferentially form intrastrand hydrogen bonds.

12 Molecule 2, since the long tract of GC pairs will require a higher temperature for dissociation of intrastrand hydrogen bonds formed after denaturation and cooling.

13 Because of the strict alternation of A and T, any region of the molecule can base-pair with any other. Thus, the molecule will fold back onto itself and form a hairpinlike structure. Of course, many different hairpinlike structures are possible, depending on the particular bases that pair. Interestingly, each structure fluctuates because of breathing, with base pairs breaking and re-forming, but not necessarily with pairing between the bases that separate. Because long tracts of AT pairs break easily at room temperature, short hairpin molecules exhibit the phenomenon of slippage—that is, the relative lengths of the two sides of the hairpin continually change as if the two parts were sliding back and forth along one another.

14 Pairing will occur randomly. Thus, there are three bands, in the ratios $1\ {}^{15}N^{15}N : 2\ {}^{14}N^{15}N : 1\ {}^{14}N^{14}N$.

15 One is T4 DNA and the other is P2 DNA; they clearly have no common base sequences.

16 Since stacking forces are a major factor in the stability of DNA, and since the terminal bases have nothing with which to stack, the terminal base pairs are frequently unbonded. The extent of loose bonding increases with increasing temperature.

17 Note that there are 14 pairs of unpaired bases in each molecule. Thus, in a molecule having 100 base pairs, 14% will be unpaired. These are terminal bases because there is less of a tendency to stack at the termini because the terminal base pair can only stack on one side.

18 There are no positive ions to neutralize the negative charges of the phosphates. The strong electrostatic repulsion causes the strands to come apart. This is a common effect in proteins also, because the repulsive electrostatic forces are greater than the attractive interactions caused by hydrogen bonding, van der Waals forces, and hydrophobic forces.

19 (2), because the effective solubility of the bases would be increased.

20 Since all parts of each strand can renature to all parts of a complementary strand, it would be unlikely that very many, if any, molecules with the original number of base pairs would form. Instead, linear structures containing an internal base-paired segment and free single-stranded ends would first form. These ends would continue to renature, forming huge linear structures containing many (thousands, millions, etc.) complementary strands. The structures would contain many single-stranded regions.

21 (a) If the ionic strength is very low, the sugar-phosphate backbone of the separated strands and all regions thereof are highly charged and mutually repulsive, thus preventing re-formation of base pairs. At high ionic strength both intra- and intermolecular base pairs form and therefore the absorbance decreases. (b) When the solution is boiled, the absorbance will increase to 1.37 but it will return to 1.00 at 25 C. (c) For the linear molecule, after returning to 25 C some interstrand base pairs will form and the absorbance will decrease somewhat. For the circular molecule, after returning to 25 C all base pairs will re-form and the absorbance will decrease to 1.00. (d) The ratio will be 1. The bases in each strand form a palindrome; that is, the sequence from bases 1 to 7 is complementary to the sequence from 14 to 8. Therefore each strand will fold back on itself and form a double-stranded hairpin.

22 (a,b,c) The second substance is a better denaturant, for it is more hydrophobic and destabilizes base-stacking more effectively. (d) The second substance has more groups that can hydrogen-bond with the bases and hence is more effective.

23 Titration of amino groups prevents hydrogen-bonding. Hydrophobic interactions maintain stacking and helicity but cannot cause base-pairing.

24 Stacking keeps the bases in line so that adjacent bases can form hydrogen bonds

with similarly oriented bases in a complementary strand.

25 In 7 M sodium perchlorate, the melting temperature must be lowered so much that at 25 C the DNA is partly denatured; in fact, it must be 48.9% hyperchromic.

26 Denatured DNA or some other ultraviolet-absorbing molecule is present in the sample. The observed value of the melting temperature would be meaningful.

27 The absorbance increase would be greater for the native DNA.

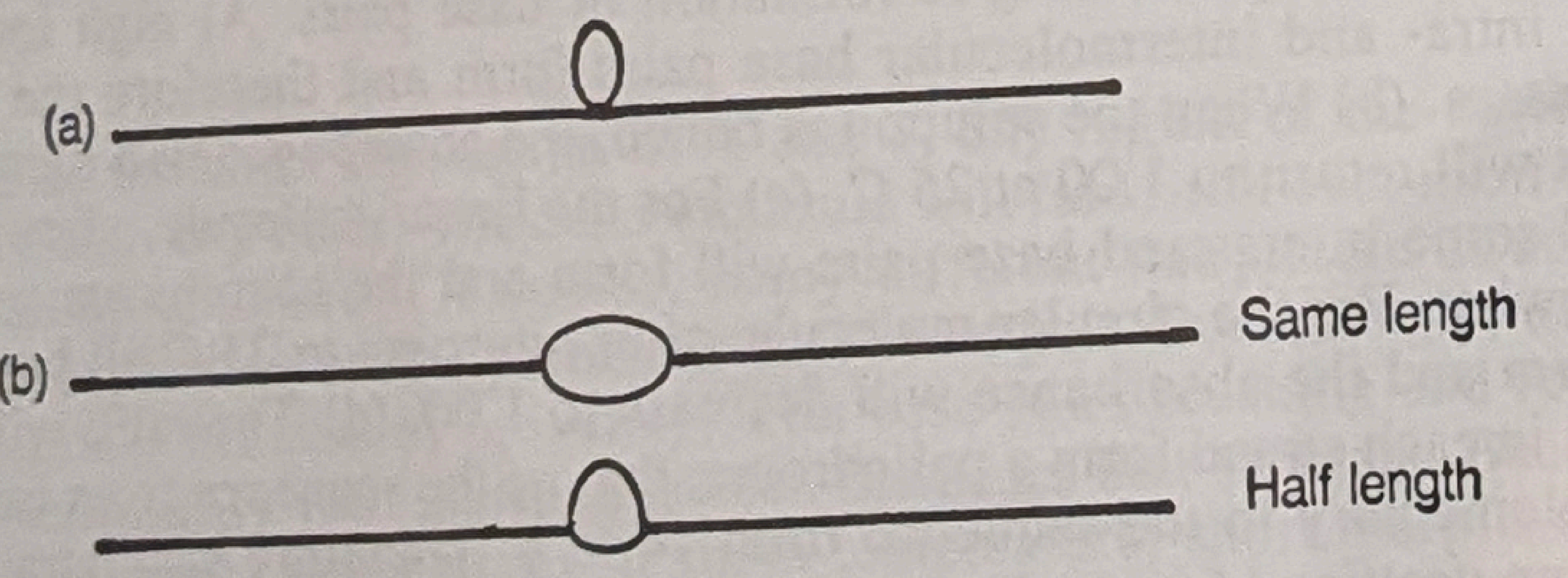
28 By shielding the negatively charged phosphates from one another the positive ions of the salt stabilizes the DNA and thus raises the melting temperature.

29 (a) The DNA is about 30% unique, 60% middle repetitive, and 10% highly repetitive. (b) The three classes have Cot midpoints of about 10^3, 10^{-1}, and 10^{-3}, respectively. Thus, the number of copies per genome is $10^3/10^3 = 1$ for the unique sequences (by assumption), $10^3/10^{-1} = 10^4$ for the middle repetitive sequences, and $10^3/10^{-3} = 10^6$ for the highly redundant sequences.

30 The mixture is 45/50 HH and 5/50 LL before denaturation. After renaturation there will be $(0.9)^2 = 0.81$ HH, $(0.1)^2 = 0.01$ LL, and $2(0.9)(0.1) = 0.18$ HL, or 40.5, 9.0 and 0.5 micrograms of HH, HL, and LL, respectively.

31 10%. Note from Problem 30 that if all sequences were common, only half would have a hybrid density.

32



33 See top of next page.

34 The sequence 5'-ATGC-3' consists of alternating purines and pyrimidines; hence, it can assume the Z-DNA form.

35 (a) 4(10) = 40 base pairs. (b) Four nodes.

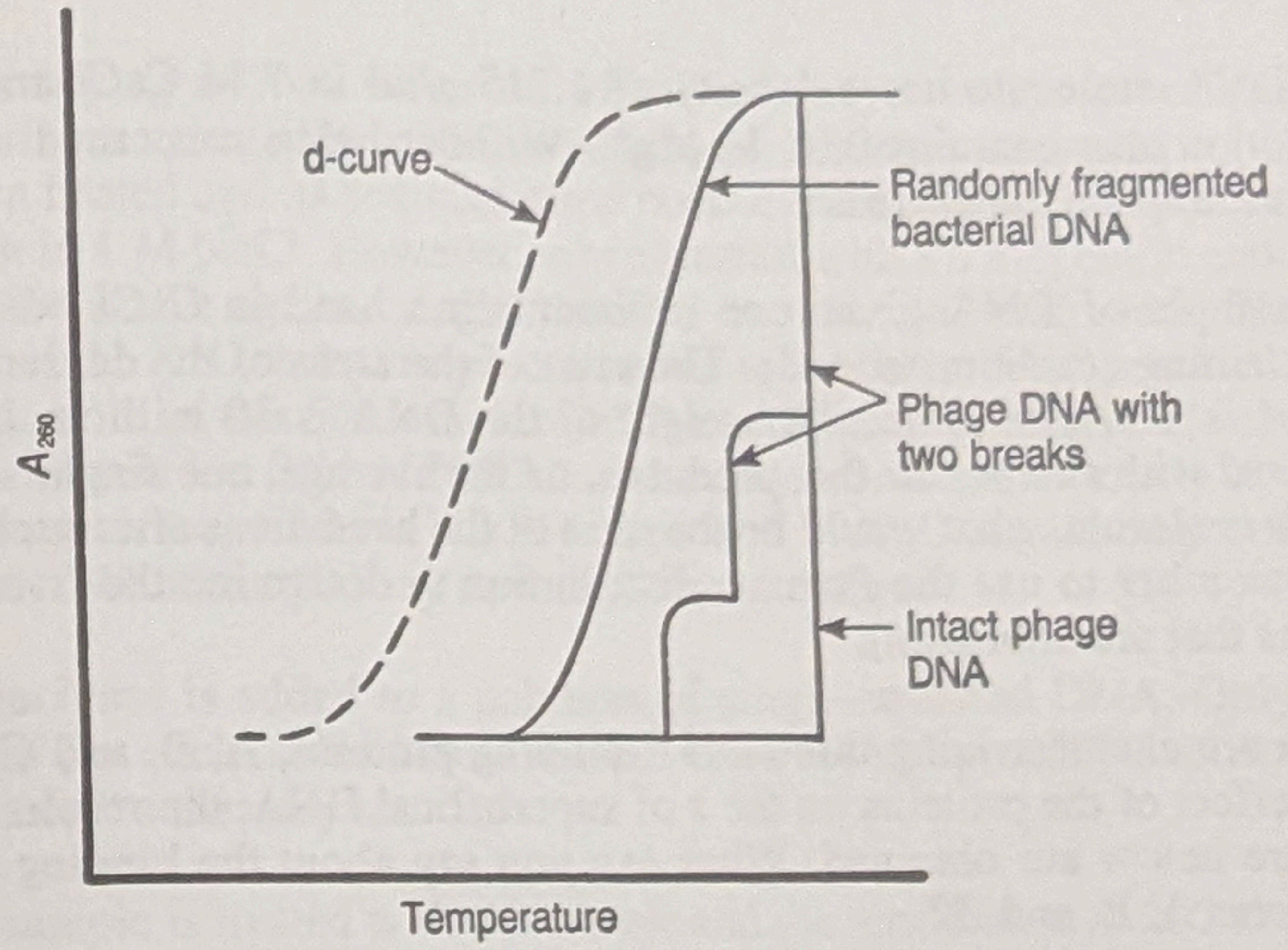


FIGURE A33

36 10 base pairs are contained in one turn of a DNA helix. Thus, the molecule will be a supercoil having one node, that is, a figure-8.

37 Each turn of the helix contains 10 base pairs, so the DNA acquires about five positive turns of supercoiling.

38 Yes, because naturally occurring supercoils are underwound and hence contain single-stranded regions.

39 The force needed is F(1/4), since a circle has half the length and twice the width of a linear molecule having the same molecular weight.

40 No, because ethidium bromide can intercalate between any two base pairs.

41 (a) (1/2)(1.592 + 1.556) = 1.574.

42 Naturally occurring DNA is negatively supercoiled. Ethidium bromide adds positive supercoiling. Thus, the compact supercoil progressively loses its supercoiling until a slower moving nontwisted circle has formed. As more ethidium bromide is bound, the DNA becomes progressively positively supercoiled and hence more compact so the value of *s* increases.

43 The enzyme fails to bind to the DNA at such high ionic strength.

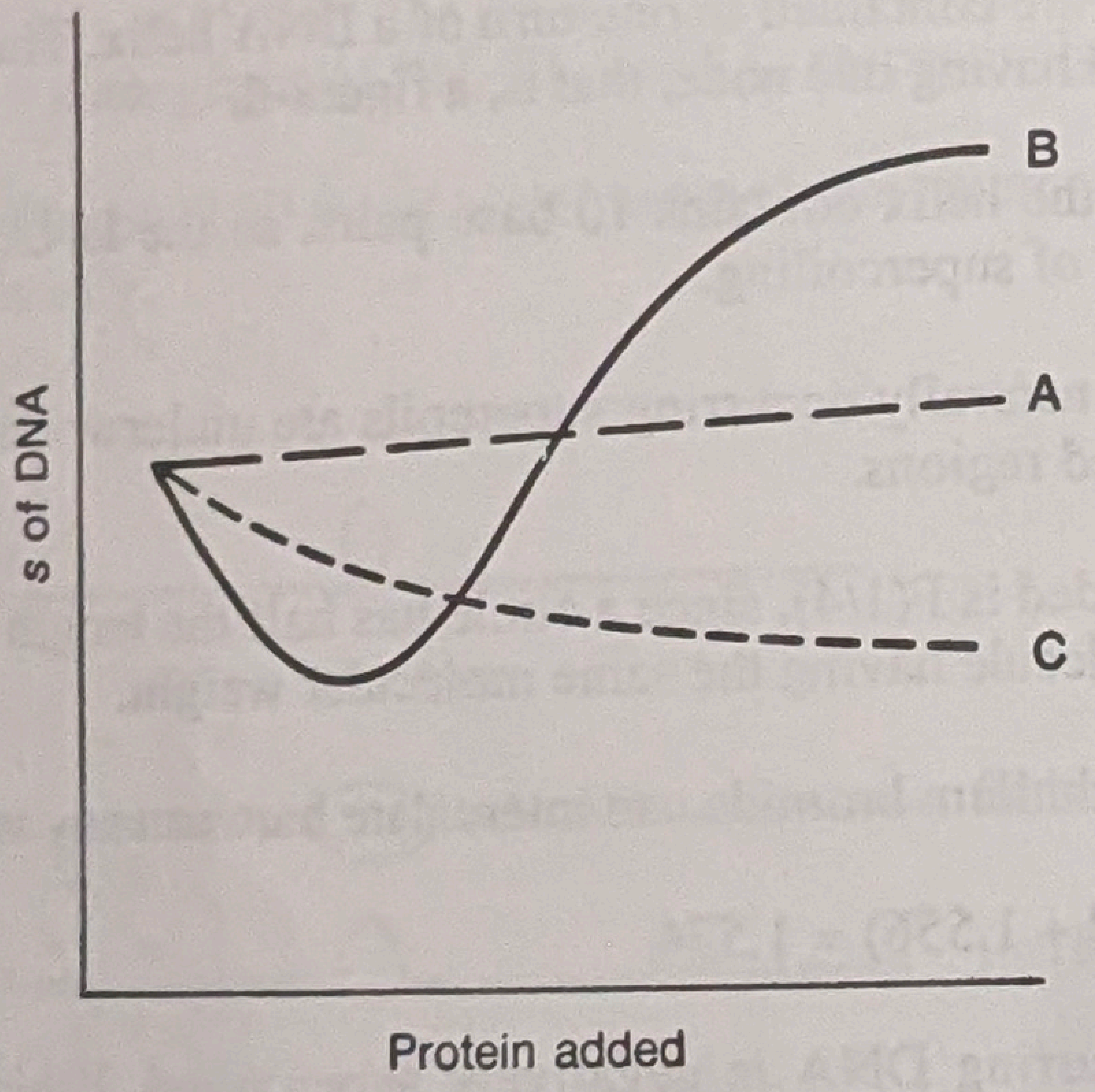
44 A few single-stranded regions are produced at temperatures before which an increase in optical density is detectable.

MORE CHALLENGING PROBLEMS

1 A DNA molecule has a density of 1.715 g/ml in 7 M CsCl and 1.699 if the solution also contains 0.01 M Mg^{2+}. Without being concerned with the exact value, explain the decrease.

2 A sample of DNA gives one sedimentation band in CsCl but two in CsCl containing ethidium bromide. The ratio of the areas of the denser to the lighter band is 2:1. The molecular weight of the DNA is 30 million. If the DNA is treated with an enzyme that produces, on the average, one single-strand break in each molecule, what would be the ratio of the band areas after such a treatment? (Remember to use the Poisson distribution to determine the fraction of molecules that are unbroken).

3 You are characterizing three DNA-binding proteins, A, B, and C, by studying the effect of the proteins on the *s* of superhelical DNA; the results shown in the figure below are obtained. What can you say about the binding properties of proteins A, B, and C?



4 In general, the viscosity ν of a solution of a macromolecule increases with both molecular weight and the ratio of length to width of the molecule. The viscosity of a DNA solution decreases with time if DNase is added, owing to the production of double-strand breaks in the DNA. Single-strand breaks have no effect on viscosity. Assume that single-strand breaks are made linearly with time during exposure to an enzyme and draw a graph of the expected change of viscosity with time for a double-stranded DNA molecule which is (a) linear, or (b) circular.

5 You are studying a DNA molecule which by many standard criteria is single-stranded: it reacts rapidly with formaldehyde, it shows no increase in absorbance when heated and its sedimentation coefficient is markedly lower in 0.01 M NaCl than in 1 M NaCl. However, when treated with a 5′P-specific exonuclease known to be active against single-stranded but not double-stranded DNA, only one-third of the DNA is hydrolyzed to mononucleotides. What possibilities exist for the structure of this DNA molecule? An additional fact learned later is that if the DNA is boiled in 0.01 M NaCl and then cooled, it can be totally hydrolyzed by the exonuclease, whereas if it is boiled in 1 M NaCl, again only one-third is hydrolyzed. What can you say now about its structure?

6 An exonuclease is added to a solution of single-stranded DNA. Only a small fraction of the nucleotides is solubilized. Give several possible explanations for the resistance to the enzyme.

7 A DNA sample is treated with a nuclease and the absorbance at 260 nm is found to increase. When the absorbance has increased by 10%, the reaction is stopped and the DNA is sedimented. The sedimentation coefficient of the DNA is only slightly different from the untreated DNA, and oligonucleotide fragments are not seen. What is a possible mode of action of this enzyme?

8 You are given a homogeneous nucleic acid sample with the following properties: (1) It reacts with formaldehyde; (2) upon denaturation its buoyant density in CsCl increases by 0.010 g/cm^3; (3) the DNA is not susceptible to either *E. coli* exonuclease I or spleen phosphodiesterase, even after alkaline phosphatase treatment; (4) its *s* is somewhat higher in high ionic strength than low ionic strength; (5) it binds to a nitrocellulose filter, which means that it probably contains single-stranded regions; (6) at low ionic strength, the *s* is too high for its molecular weight, if it were simply double-stranded DNA; (7) centrifugation of the denatured material gives rise to two bands, A and B. Exposure of purified B for a short period of time to pancreatic DNase results in a molecule C with a single break. The *s* of C is about the same as B at high ionic strength, but lower than B at low ionic strength. The *s* of C is higher than A at low and high ionic strength. What is the structure of the nucleic acid?

9 A phage mutant has been isolated whose DNA has the following properties: (1) DNA isolated from the phage sediments as a single component (structure A) with a sedimentation coefficient that is relatively independent of the ionic strength; (2) molecule A does not react with formaldehyde, and acridine compounds cause the viscosity to increase and the sedimentation coefficient to decrease; (3) sedimentation at pH 12.5 reveals two sedimenting components and twice as much material (measured by absorbance) sediments in the faster component than the slower component; (4) when A is heated to 10 C below the melting temperature and cooled rapidly, two components B and C are produced, which have different sedimentation coefficients, and when cooled slowly, only

one sedimenting component is seen and it has the same sedimentation coefficient as A; (5) limited exposure of A to exonuclease III, followed by annealing, results in a structure D that has a higher *s* than A; (6) treatment of D with DNA ligase (which seals single-strand breaks) produces E, which in alkali sediments as a single component at least four times faster than the larger, denatured component of A; (7) when poly(G) is added to heat-denatured A, two bands differing in buoyant density are seen in CsCl density gradients; (8) if A is treated with DNA ligase, which joins a 5'-P group and a 3'-OH group, then terminally labeled with ^{32}P by the polynucleotide kinase reaction, and then digested to mononucleotides, two radioactive nucleotides, dAMP and dGMP, are found. If after the labeling but before digestion the strands are fractionated as in (8) above, the labeled dGMP can be shown to be associated with the strand that binds more poly(G). What is the structure of the phage DNA?

10 A nucleic acid molecule having a G+C content of 50% has the following properties: (1) The denaturation temperature and buoyant density in CsCl are higher than that of *E. coli* DNA; (2) the molecule is not cleaved by RNase; (c) a single scission by DNase gives rise to structure A, having a lower sedimentation coefficient than the untreated molecule; (4) denatured A (which is called B) gives rise to two bands when centrifuged to equilibrium in neutral CsCl, but only one in alkaline CsCl; (5) exposure of the original molecule to pH 13 for 60 minutes at 37 C and reneutralization gives rise to a structure C which has a lower buoyant density in neutral CsCl than before alkaline treatment; (6) molecule C is resistant to snake venom phosphodiesterase; (7) when C plus the pH 13-digestion products are filtered through a nitrocellulose membrane filter, half of the material, which is D, passes through the filter, and the material retained on the filter can be eluted and is found to be C; (8) polynucleotide kinase does not phosphorylate either forms C or D. What are the structures of the parent molecule and of its derivatives?

ANSWERS TO MORE CHALLENGING PROBLEMS

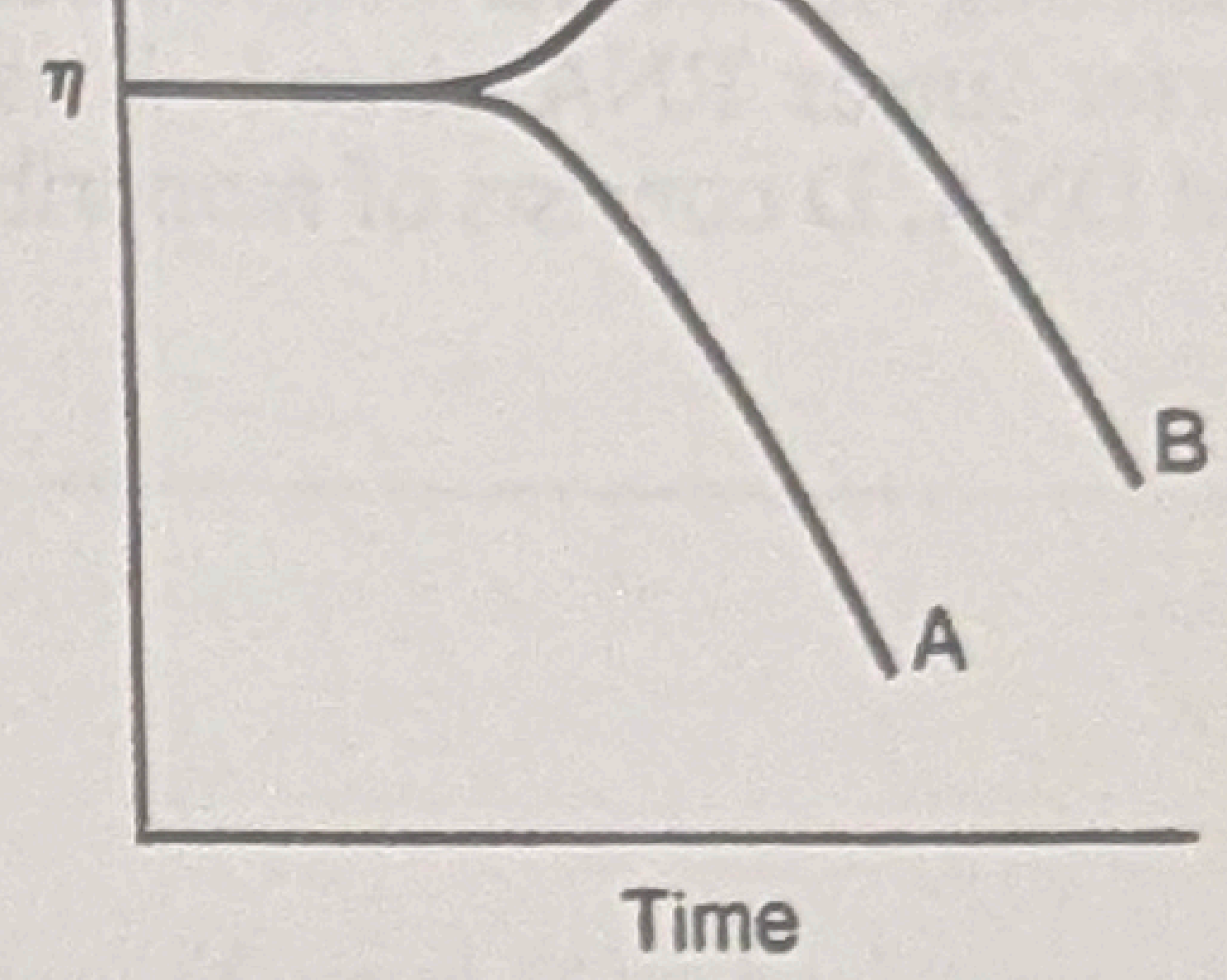
1 In CsCl the DNA has the density of the Cs salt of DNA. Since the Mg^{2+} ion binds tightly to DNA, even in CsCl, the DNA is in the form of MgDNA, which has a lower density because of the lower density of the Mg ion compared to the Cs ion.

2 The denser band is clearly a supercoil. When there is an average of one break per molecule, the fraction unbroken is 1/e or 37%. Therefore, the 2:1 ratio would become 2(0.37):[1 + 2(1 - 0.37)], or 0.74:2.26, or approximately 1:3.

3 Protein A may bind very weakly to superhelical DNA and may possibly introduce superhelical twists of the same sense as those in the DNA molecule. Alternatively, it binds strongly but does not alter *s* other than by increasing the molecular weight. Protein C may bind weakly, and thereby remove some superhelical twists, but it may also be a highly asymmetric molecule that increases the frictional coefficient of the DNA molecule decreasing *s* (by

cancelling the effect of superhelicity). Protein B introduces superhelical twists in the opposite sense from those already in the DNA.

4



5 The DNA molecule must contain something other than single-stranded DNA one-third of the way from the 5'-P terminus. Possibilities include protein or RNA linkages within the polynucleotide chain, a protein molecule bound to a sequence of DNA bases, a short sequence of double-stranded DNA or RNA, or a hairpin. The boiling experiment indicates reversible thermal disruption at high but not low ionic strength; this strongly implicates the hairpin.

6 Possibilities include a circular DNA molecule, altered termini (for example, altered 3'-P or 5'-OH groups), a blocked terminus (for example, blocked by an ester), and a hairpin producing a double-stranded terminus.

7 It is probably an exonuclease acting on only one type of terminus (3' or 5'). Thus, single stranded DNA is produced at both ends. Both the single-stranded regions and the mononucleotides have a higher value of the absorbance than double-stranded DNA, which accounts for the absorbance increase. The single-stranded ends are flexible and compensate for the decrease in molecular weight of the entire molecule in determining the *s* value.

8 Results 1, 4, and 5 show that there is some single-stranded DNA. Result 2 shows that there is some double-stranded DNA; the shift of 0.010 g/cm^3 rather than 0.014 g/cm^3, suggests that about 70% of the DNA is double-stranded. Result 3 suggests that the molecule is circular; this is confirmed by result 6. Result 7 shows that the two component strands do not have the same length. The molecule consists of one circular strand to which is hydrogen-bonded a linear strand whose molecular weight is roughly 70% that of its circular strand.

9 The DNA is double-stranded without single-stranded termini but with a single-strand break in each strand. These strand breaks are each located 1/3 of the

distance from the 5' end. The strand binding poly(G) has guanine at the 5' terminus; the other strand has adenine.

10 The high buoyant density suggests that RNA is present. Insensitivity to RNase shows that the RNA is not single-stranded. Result 1 shows that C is single-stranded. Result 6 shows the DNA is circular. Results 3 and 4 shows that the two strands are not the same length. The single band in alkaline CsCl and result 5 are caused by alkaline degradation of RNA to mononucleotides. A is a circular molecule consisting of circular single-stranded DNA to which is hydrogen-bonded a shorter linear RNA strand. B is a linear form of A. C is circular single-stranded DNA. D consists of monoribonucleotides.

CHAPTER 5

DNA TECHNOLOGY

CHAPTER SUMMARY

DNA can be isolated in a variety of ways. Two basic steps are required: breaking open of the organism containing the DNA, and separation of the DNA from other cellular components. Isolation of DNA from phages and viruses is straightforward: treatment with phenol both releases the DNA and precipitates protein. For bacteria and fungi an enzymatic treatment is needed to break open the cell (lysozyme for bacteria, a snail enzyme for fungi); phenol is used to precipitate protein and to remove many organic molecules that are more soluble in phenol than in water. A final step in which DNA is precipitated with ethanol provides additional purification. With animal cells nuclei are usually isolated, which reduces contamination by protein and RNA and enriches for DNA, and the phenol and ethanol treatments are then used.

Nucleic acids can be radioactively labeled either by growth of organisms in radioactive medium or by enzymatic labeling of isolated material. Radioactive DNA can be prepared by growing cells in medium containing [^{3}H]thymidine. Only DNA becomes labeled. ^{32}P is also useful but DNA, RNA, and phospholipids must be separated. Radioactive RNA can be prepared by using any radioactive base or ribonucleoside, but separation of RNA and DNA is always necessary. DNA can be terminally labeled at the 5' end by a reaction with labeled ATP catalyzed by polynucleotide kinase. The terminal 5'-P must first be removed with alkaline phosphatase because the kinase is active only on 5'-OH groups.

Restriction enzymes are site-specific endonucleases that make two cuts within specific base sequences. Each enzyme recognizes only a single sequence, and the sequence is always a palindrome, ranging from 4 to 6 nucleotide pairs. Most restriction enzymes make two cuts, one in each strand, separated by a few bases, and hence produce complementary single-stranded termini. Some enzymes make two central cuts directly opposite one another, generating blunt ends. Because of a limited number of target sequences for each enzyme, the number of cuts made by a particular restriction enzyme is a small fraction of the total number of phosphodiester bonds. This feature enables one to characterize a particular DNA molecule by a restriction map, a map showing the sites of cutting by a single enzyme. A variety of methods, of which the double-digest procedure is generally the most useful, can identify the sites of cutting.

Many bacteria contain small replicating circular DNA molecules called plasmids. Plasmid DNA is usually separated from the bulk of the bacterial DNA by the cleared-lysate procedure, in which the bacterial chromosome is precipitated as a protein-detergent complex. Plasmid DNA, because of its small size, usually remains in the supernatant. Final purification of plasmid DNA makes use of the fact that it is a covalently closed circle and uses equilibrium centrifugation in CsCl containing the intercalating agent ethidium bromide. This substance lowers the density of DNA. For steric reasons, covalent circles bind less ethidium bromide than linear molecules and open circles, so their density is shifted less than these other forms.

Since a particular restriction enzyme produces a unique pair of complementary termini at the sites of cutting, molecular fragments cleaved from different DNA molecules by a single enzyme can be joined by annealing. If a fragment is joined to a replicable molecule such as a plasmid (which has been cleaved by the same enzyme), the fragment can be cloned. Plasmid DNA can be taken up by *E. coli* and several other bacteria by exposing the cells to solutions of $CaCl_2$. The plasmid DNA, which may contain an inserted restriction fragment, can become established as a stable replicating element of the bacterium, thereby carrying the cloned fragment into all daughter cells.

The complementary strands of a DNA molecule can frequently be purified by mixing denatured DNA with a ribopolymer. The basis of the technique is the existence of unequal numbers of either purine or pyrimidine tracts on the complementary strands. Thus, hybridization with a polypurine will cause more of the ribopolymer to be bound to one strand than to the other. Since ribopolymers have a higher density in Cs salts than DNA, the two complementary strands, by virtue of the amount of bound ribopolymer, will have different densities. Hence they can be separated by equilibrium centrifugation in Cs salts.

DNA-RNA hybridization can be used to identify particular restriction fractions. The Southern blotting technique enables one to perform a large number of hybridizations without purification of the individual fragments. A gel containing restriction fragments is treated with alkali to denature the DNA and then placed in contact with a sheet of nitrocellulose; the denatured fragments are transferred from the gel to the sheet. A hybridization probe, which may be radioactive single-stranded DNA or RNA, is then added and renaturation is allowed to occur. Radioactive positions are located by autoradiography.

The nucleotide sequence of a DNA fragment can be determined by the Maxam-Gilbert procedure. In this method a terminally labeled single-stranded DNA molecule is treated in such a way that four families of fragments are produced, one family for each DNA base. The sizes of the members of each family are determined by the location of the particular base measured from the 5' end. Fragment size is determined by gel electrophoresis. The presence of a fragment with *n* bases in a family that locates thymine, for example, indicates that a thymine is present in the original molecule at position $n+1$ from the 5' terminus.

Double-stranded DNA, one of whose strands is complementary to an RNA molecule, can be prepared by exposure of the RNA to reverse transcriptase, an enzyme isolated from certain animal viruses. In the reaction carried out in the laboratory reverse transcriptase is used to convert RNA to a DNA-RNA hybrid. The RNA is then removed by a specific RNase, and the single-stranded DNA is used as a template for synthesis of double-stranded DNA, typically using a DNA polymerase from *E. coli* called polymerase I.

BOLD TERMS

blotting, blunt ends, c-DNA, $CaCl_2$ transformation, cell lysate, cleared lysate, cloned, cohesive ends, complementary DNA, double-digest procedure, flush ends, lysozyme, Maxam-Gilbert procedure, Northern blotting, partial digest, plasmid, polynucleotide kinase, primer, probe, proteases, restriction endonuclease, restriction map, reverse transcriptase, Sanger dideoxy procedure, S1 nuclease, Southern transfer procedure, Western blotting

DRILL QUESTIONS

1 What is the function of phenol in many procedures for isolating DNA?

2 How is DNA usually collected from a solution containing small amounts of other molecules?

3 How is RNA removed when DNA is to be isolated from cellular material?

4 How is DNA removed when RNA is to be collected?

5 Can DNA be labeled with ^{32}P using labeled ATP and polynucleotide kinase as the only enzyme?

6 Why can radioactive thymidine be used as a specific radioactive label for DNA?

7 Why can uridine, which is not present in DNA, not be used as a specific label for RNA?

8 Restriction enzymes generate three types of termini. What are they?

9 Are the termini of a restriction fragment produced by a particular enzyme always the same or can they be different?

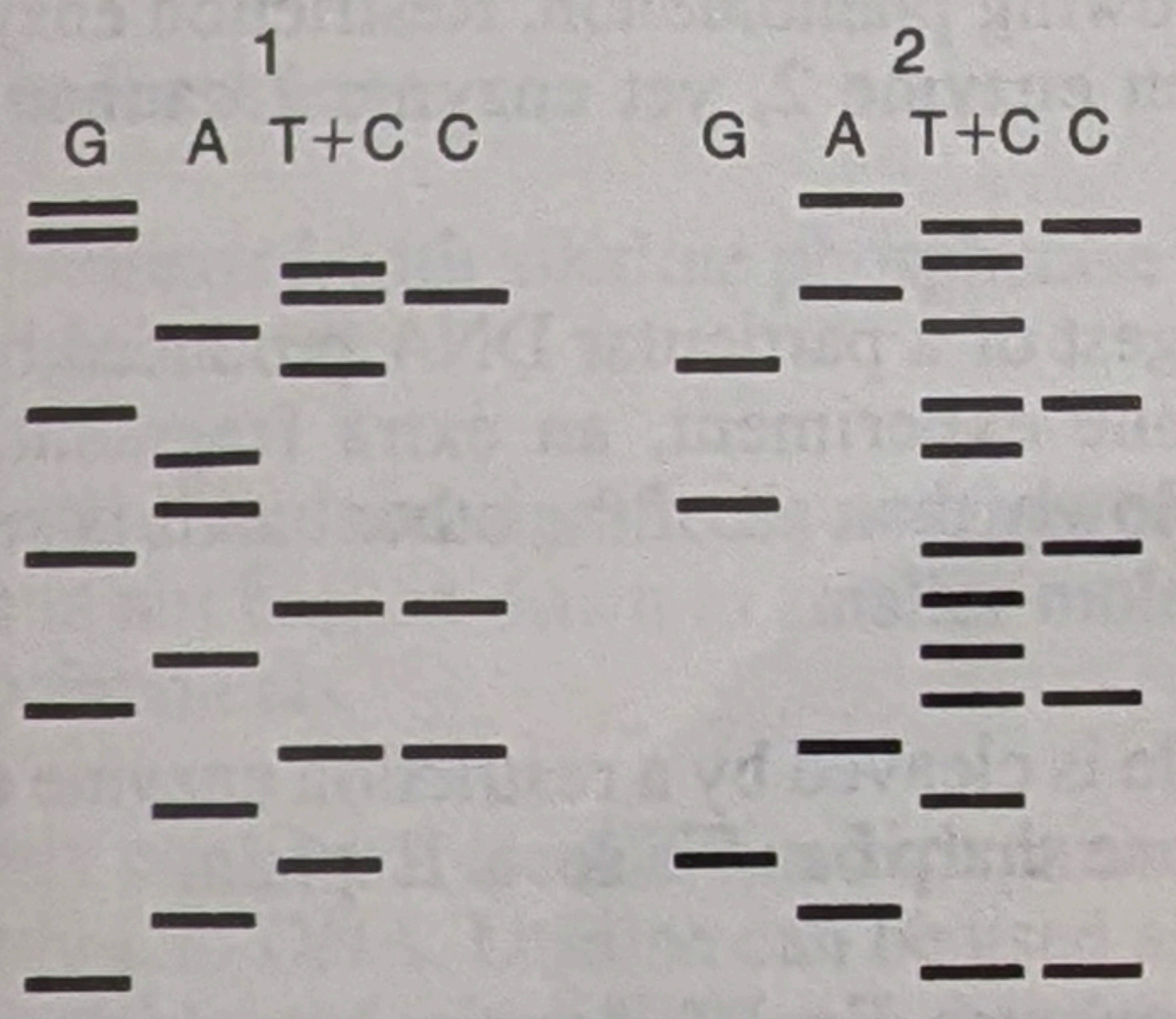
10 Can a restriction enzyme cut at two sites having different base sequences?

fragments are produced, whose sizes are 1.3, 2.8, 3.6, 5.3, 7.4, 7.6, 8.1, and 11.4 kb. However, if J2 DNA is isolated from infected cells, only seven fragments are found, whose sizes are 1.3, 2.8, 7.4, 7.6, 8.1, 8.9, and 11.4. What is the likely form of the intracellular DNA?

7 If a plasmid is cleaved in five positions with a restriction enzyme that produces an extended 3' terminus, will all fragments be able to circularize?

8 Phage XQ42 DNA is cleaved into six fragments by the enzyme BglI. A mutant is isolated whose plaques look quite different from the wild-type plaque. DNA isolated from the mutant is cleaved into only five fragments. Account for this change. In general, discuss why often a restriction enzyme may be found that yields *n* fragments with wild-type phage and *n* - 1 fragments with a mutant.

9 A DNA fragment containing 17 base pairs is sequenced by the Maxam-Gilbert procedure. The figure below shows the data; panels 1 and 2 correspond to the two complementary strands. Note that the 5'-terminal base does not appear in either lane, as it would have to be identified by a sugar-phosphate lacking a base; such molecules do not electrophorese with the nucleotides. What are the complete sequences of the two complementary strands, including the 5'-terminal bases?



10 The nucleotide sequence of a region of a single-stranded DNA molecule is determined by the Maxam-Gilbert procedure. Data for a particular sample are shown in Figure 10. (a) What is the sequence of the two DNA samples shown in panels (i) and (ii)? (b) A DNA molecule much longer than those in (b) can be sequenced. However, often all of the bands cannot be resolved on a single gel. To get around this, the reacted samples are divided into two parts, and each is treated by electrophoresis but for different lengths of time. What is the sequence of the DNA shown in panel (iii)?

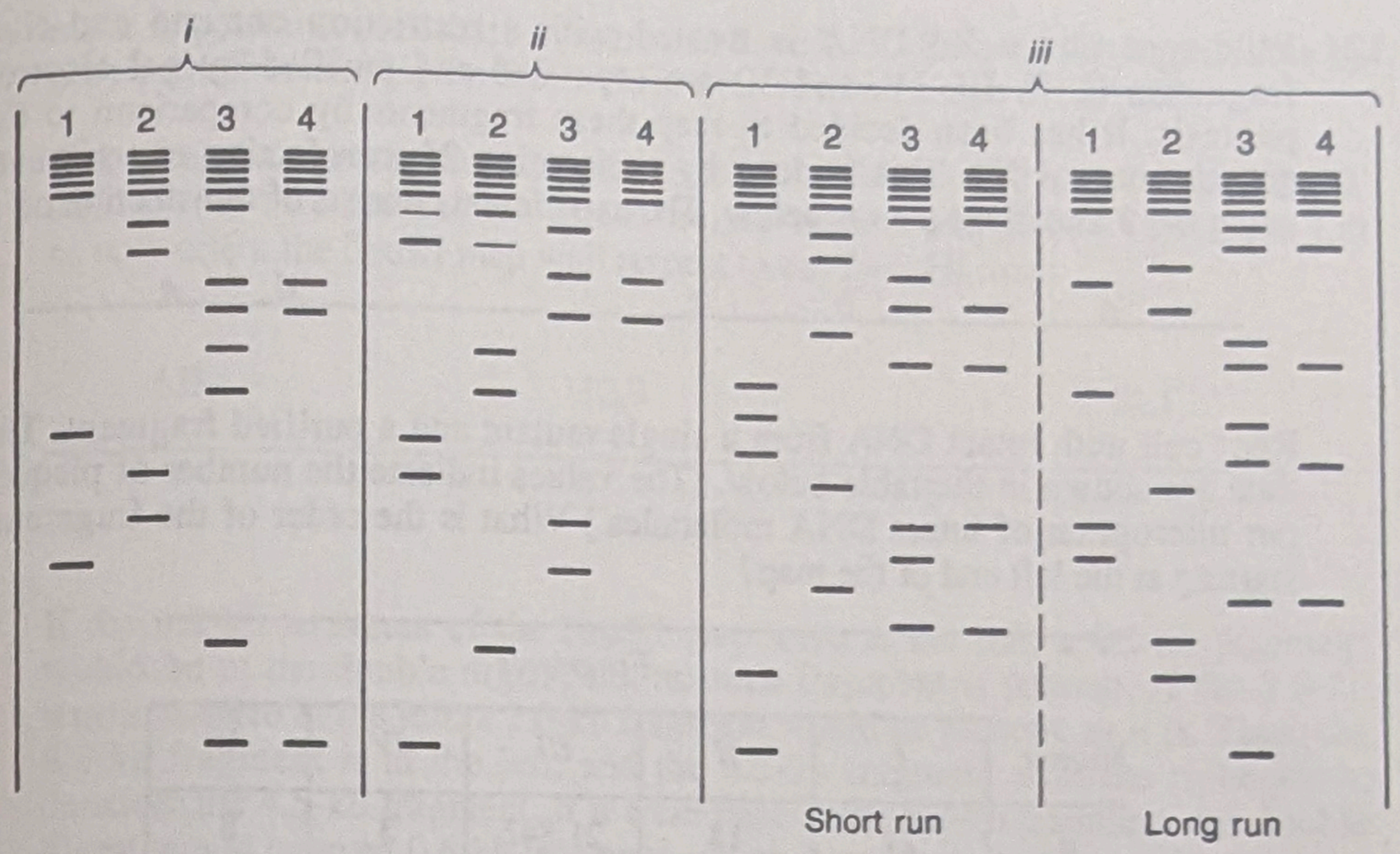
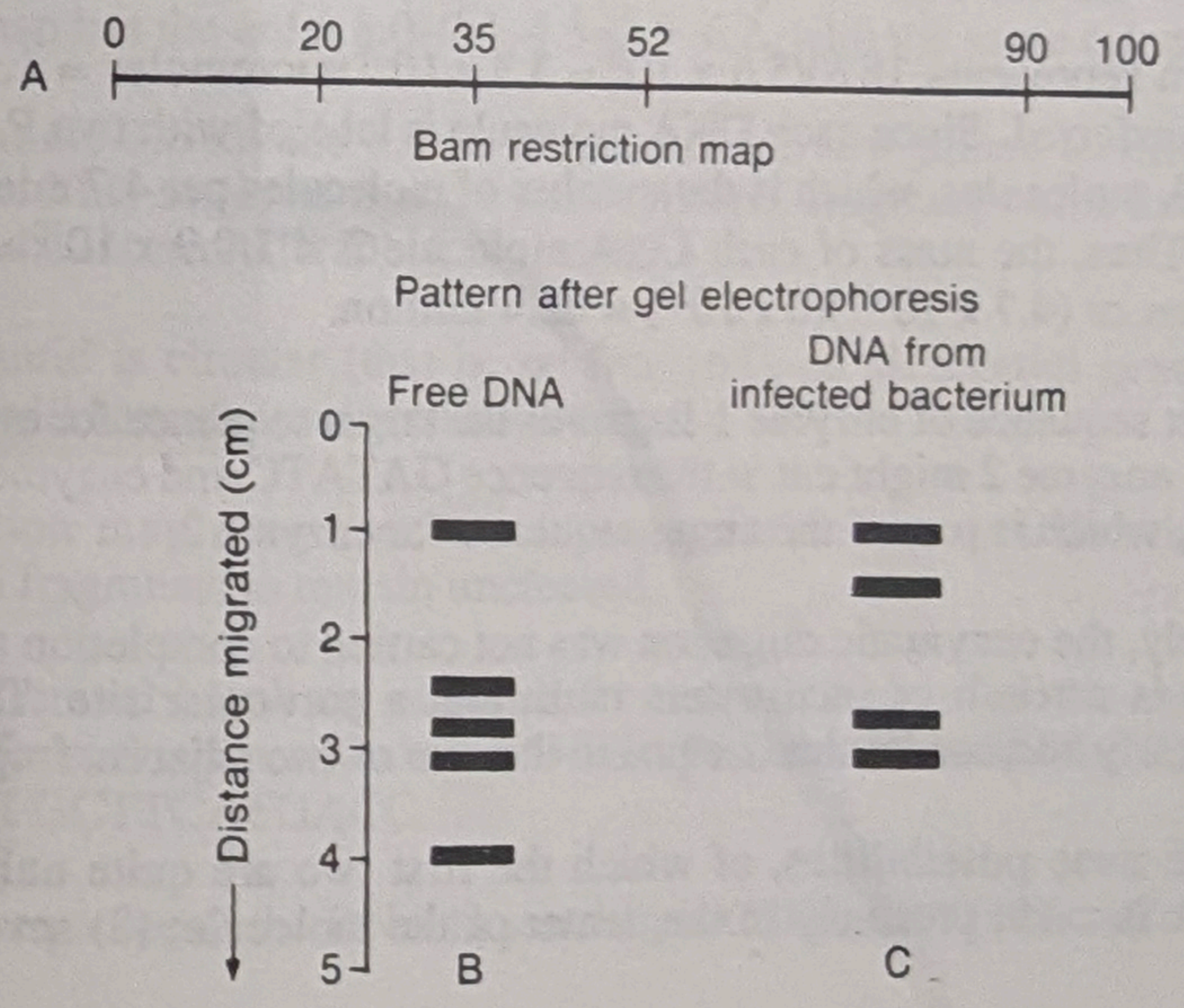
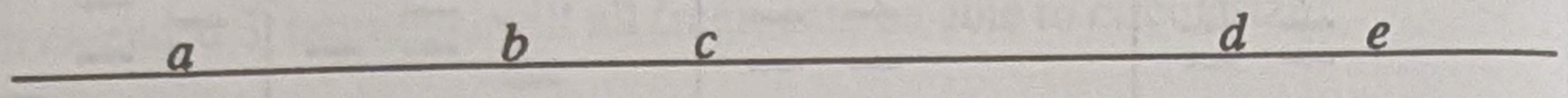


FIGURE 10

11 A phage DNA molecule has short, complementary, single-stranded ends and circularizes when the phage infects a bacterium. The restriction map and the gel band pattern obtained after treatment of a free DNA molecule with the BamHI restriction enzyme are shown at A and B in the figure below. Enzymatic digestion of DNA isolated from infected bacteria yields the gel band patterns shown in lane C. What is the structure of the DNA under these conditions?



12 Wild-type phage X4 DNA is treated with a restriction enzyme and five fragments (I, II, III, IV, and V) are separated and purified by gel electrophoresis. It has been decided to map these fragments by comparison to the genetic map of X4. This is done by performing 25 transfection experiments using the 5 mutations shown below. The experiment consists of transfection of a



Rec$^+$ cell with intact DNA from a single mutant and a purified fragment. The data are shown in the table below. (The values indicate the number of plaques per microgram of intact DNA molecules.) What is the order of the fragments starting at the left end of the map?

	Fragments				
Mutant	*I*	*II*	*III*	*IV*	*V*
a^-	21	14	21,842	3	8
b^-	6	0	4	4,856	2
c^-	32,681	16	8	11	13
d^-	1	3	0	1	1,825
e^-	22	17,813	27	18	30

ANSWERS TO PROBLEMS

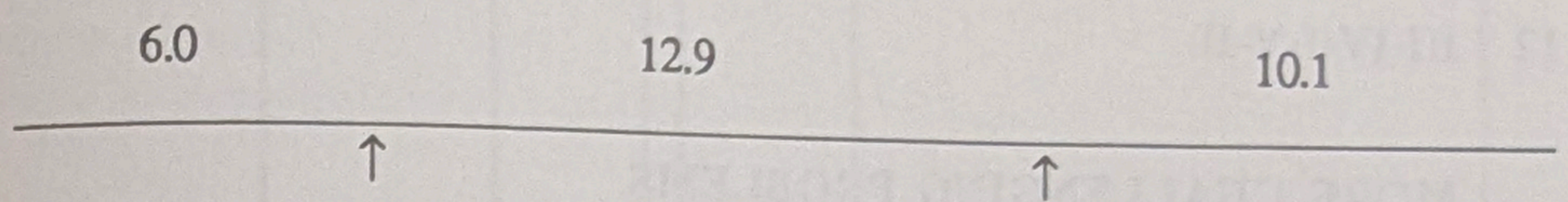
1 1870 cpm represents $1870/5.6 \times 10^9 = 3.3 \times 10^{-7}$ micromoles $= 1.98 \times 10^{11}$ P atoms transferred. Since each DNA molecule is labeled with two P atoms, 9.9×10^{10} DNA molecules, which is the number of molecules per 4.7 microgram , are labeled. Thus, the mass of each DNA molecule is $4.7/9.9 \times 10^{10} = 4.7 \times 10^{-11}$ microgram or $(4.7 \times 10^{-11})(6 \times 10^{17}) = 22.4$ million.

2 The target sequence of enzyme 1 is part of the target sequence for enzyme 2. For example, enzyme 2 might cut in the sequence GATATC and enzyme 1 might act at ATAT, which is part of the target sequence for enzyme 2.

3 Most likely, the enzymatic digestion was not carried to completion and the extra fragment is a result of incomplete cutting at a particular site. The fragment moves slowly because its size is equal to the sum of two adjacent fragments.

4 There are three possibilities, of which the first two are quite unlikely: (1) A single cut is made precisely in the center of the molecule; (2) several cuts are made in positions such that all fragments have sizes that are indistinguishable by gel electrophoresis; (3) the molecule is a circle and a single cut is made.

5 The labeled fragments indicate the termini, which are 6.2 and 8.0 for EcoRI, and 6.0 and 10.1 for BamHI. Therefore, the BamHI map is 6.0, 12.9, 10.1. Let us now orient the EcoRI map with respect to the Bam HI map:



If the 6.2-kb terminus of the EcoRI map were at the left, a 0.2-kb fragment would be in the double digest, but no such fragment is present. If the 8.0-kb terminus were at the left, a 2.0-kb fragment would be present, as it is. Thus, the 8.0-kb fragment is at the left, and the 6.2-kb fragment is at the right. Now, consider the 4.5-kb fragment. If it were next to the 6.2-kb fragment, the double digest would have an 0.6-kb fragment, which is not present. Also, if it were next to the 8.0-kb fragment, the double digest would have a 6.5-kb fragment, which it does not. Therefore, the 4.5-kb fragment cannot be next to either the 6.2-kb or the 8,0-kb fragment and must be in the center of the molecule. Now, consider the 7.4-kb fragment. If it were next to the 8.0-kb fragment, there would be a 2.5-kb fragment, which there is. If it were instead next to the 6.2-kb fragment, there would be a 3.5-kb fragment, which is not present. Thus, the 7.4-kb and 6.2-kb fragments are adjacent. Analysis of the position of the 2.9-kb fragment shows that it is next to the 8.0-kb fragment, which agrees with the position of the 7.4-kb fragment. Therefore, if the BamHI map has the order 6.0–12.9–10.1, the EcoRI map has the order 8.0–7.4–4.5–2.9–6.2, with the same orientation.

6 Note that the 3.6-kb and 5.3-kb fragments are now joined to form the 8.9-kb fragment. This suggests that they are terminal fragments of a linear molecule and that the intracellular DNA is circular.

7 The plasmid is circular (that is, no free end), so all termini generated by the enzyme will have the same cohesive ends.

8 A mutation may alter one base in a restriction site and thereby cause two potential fragments to remain uncleaved.

9 The 5'-terminal bases are obtained from the 3' termini of the complementary strand. Thus, the two strands are 5'-TGATACGACGAAGTACTGG and 3'-AC TATGCTGCTTCATGACC.

10 (a) The sequence of the DNA in panel (i) is CTGAAGTTCCAAT. That of panel

(ii) is GATTGGAACCGTA. (b) The short run gives the sequence GGC ATCAGGGCACTAAC. The long run gives the sequence TAACGGACTGC TAGACT. Presumably the TAAC is common to both so the sequence is GGCAT CAGGGCACTAACGGACTGCTAGACT.

11 The terminal fragments are missing and a new fragment is present whose size is the sum of the sizes of the missing fragments. Thus, the DNA is circular.

12 III-IV-I-V-II.

MORE CHALLENGING PROBLEMS

1 The following elegant method has been used to obtain the restriction maps of a DNA molecule for different enzymes. First, the 5' termini of a phage are labeled with ^{32}P by the polynucleotide kinase reaction. The sample is then divided into two parts and each is digested either with enzyme I or enzyme II. The two digests are separately electrophoresed and the positions of the terminal fragments are identified by the radioactive bands. The major part of the technique uses both an unlabeled and a uniformly ^{32}P-labeled sample of the DNA. The unlabeled sample is digested with enzyme I and then electrophoresed. These bands are located and their positions are recorded. The DNA is then denatured by immersing the gel in alkali, and the bands are transferred by Southern blotting to a nitrocellulose sheet; this yields the configuration shown in Figure C1. The ^{32}P-labeled sample is treated with enzyme II and electrophoresed and the positions of the bands are noted. The DNA is denatured with alkali, and then transferred to the nitrocellulose paper containing the enzyme I fragments. The transfer is done such that the ^{32}P bands are at right angles to the unlabeled bands, as shown in the figure. The paper is then subjected to renaturing conditions for several hours. It is then washed, and the radioactivity is located by autoradiography. The figure shows the positions of the bands from each gel (thin lines) and the positions of the radioactivity (small circles). The terminal fragments are indicated by asterisks. What are the restriction maps for each enzyme? The numbers indicate the size of the fragment in kilobase pairs; the solid circles indicate radioactive regions.

2 The DNA of phage P4 is linear, double-stranded, 11.5-kb long, and has complementary single-stranded ends. Digestion with the BamHI restriction endonuclease yields fragments 6.4, 4.1 and 1.0 kb in length. Partial digestion with the BamHI enzyme yields fragments 10.5, 7.4, 6.4, 4.1, and 1.0 kb in length. Circular P4 DNA made by joining of the complementary termini and enzymatically converted to a covalent circle by sealing the single-strand interruptions can be digested with the BamHI enzyme to yield fragments 6.4 and 5.1 kb in length. What is the order of fragments in the DNA?

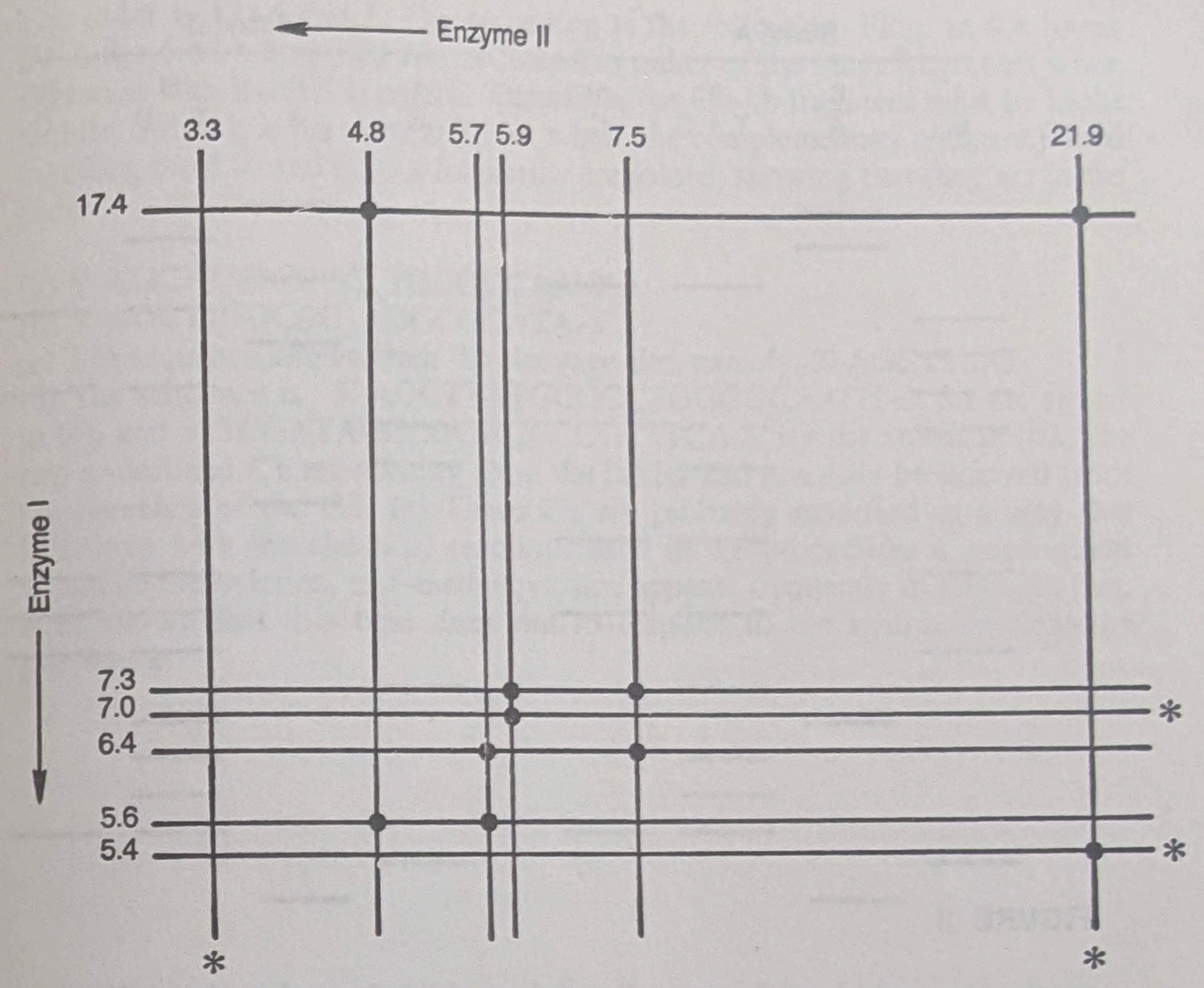


FIGURE C1

3 A DNA fragment is isolated by cutting a large DNA molecule with the HindIII restriction enzyme, which makes cuts in a particular base sequence at positions show by the arrows in the figure below.

↓
Hind III . . . pTpTpCpGpApA-5′
5′-ApApGpCpTpTp . . .
↑

The sequence of each strand is determined by labeling the 5′ ends with ^{32}P (using the polynucleotide kinase reaction) and then using the Maxam-Gilbert sequencing method. The mixture of cleaved fragments is electrophoresed yielding the data of Figure 3 (following page). Observe the spacings in the ladder very carefully and answer the following questions. (a) What is the sequence of strand A? (b) What is the sequence of strand B? (c) What part of the

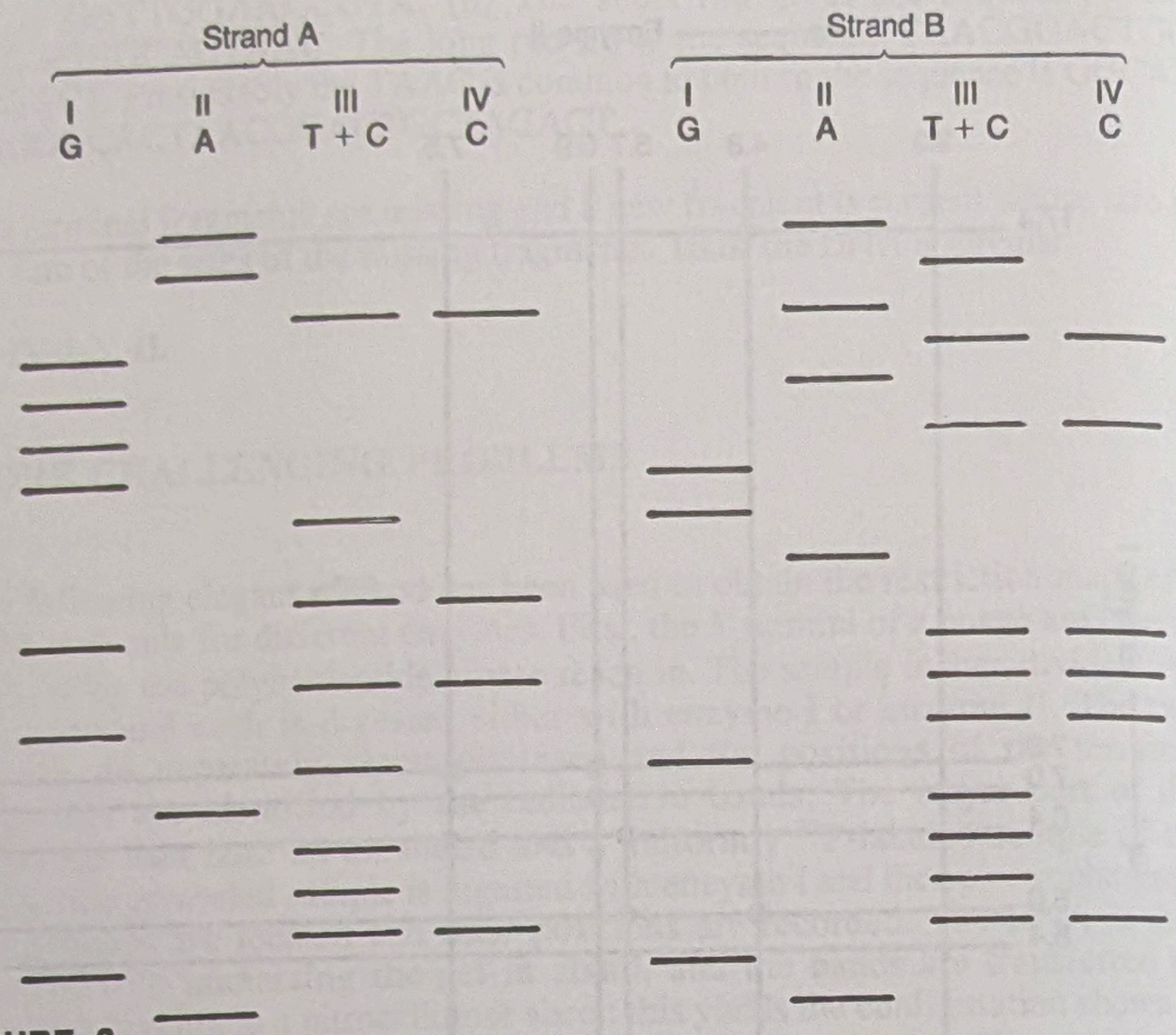
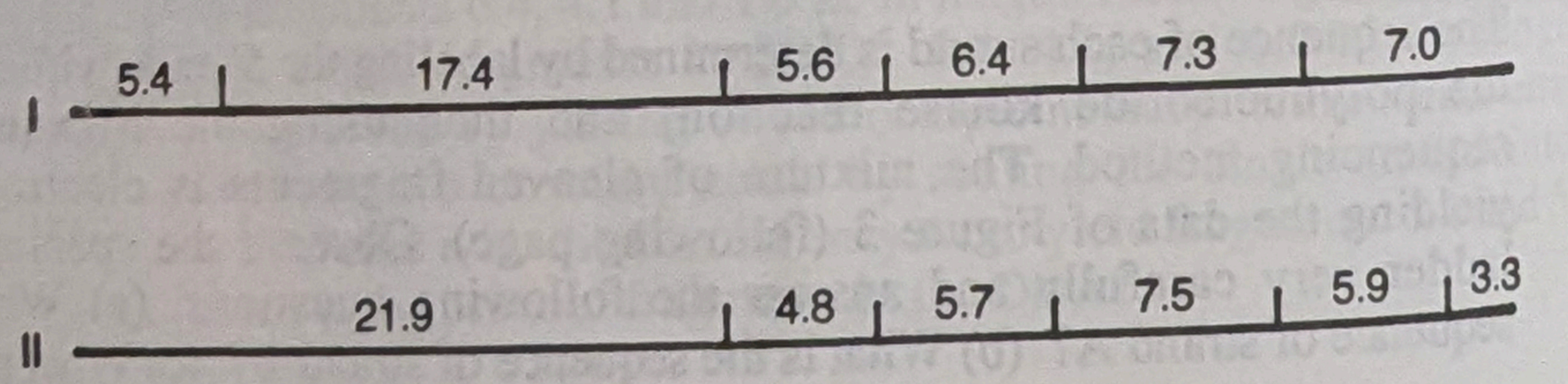


FIGURE 3

sequences could have been predicted from the information about the sites of cleavage? (d) Line up the complementary parts of the two strands as best you can. What irregularity can you see? (e) From what you may know about unusual bases occasionally observed in DNA, suggest an explanation for the irregularity in part (d).

ANSWERS TO MORE CHALLENGING PROBLEMS

1 The principle underlying the technique is that ^{32}P is found at intersections of overlapping fragments.



2 The order is 1.0–6.4–4.1. The reasoning is the following. First, in the linear DNA the 6.4-kb fragment can be linked to either of the other fragments when digestion with BamHI is partial. Therefore, the 6.4-kb fragment must be in the middle. Second, in the circular DNA, where the complementary ends are joined together, the 1.0- and 4.1-kb fragments are joined, showing that they are at the ends.

3 (a) 5'-AGCTTATGCGC_TGGGGCAA-3'.
(b) 5'-AGCTTTGCCC_AGGCGCATA-3'.
(c) The sequence known from the cleavage site, namely, 5'-AGCTT...-3'.
(d) The sequence is 5'-AGCTTATGCGCCTGGGGCAAGT-3' for the strand in (a), and 3'-TCGATACGCGGACCCCGTTTCA-5' for the strand in (b). The two underlined C's are missing from the ladder and can only be inferred from the presence of the G's. (e) These C's are probably modified in a way that interferes with the chemical reactions used in the procedure. A good guess would be methylation, as 5-methylcytosine appears frequently in DNA. In fact, it is known that this base does not to respond to the hydrazine-cleavage procedure.

CHAPTER 6

PHYSICAL STRUCTURE OF PROTEIN MOLECULES

CHAPTER SUMMARY

Proteins are linear polymers of amino acids, ranging in size from about 10 amino acid residues for the smallest chains to 1500 for the largest. The majority of proteins contain several hundred amino acids and have molecular weights of tens of thousands. Polypeptide chains are formed by a reaction between the α-carboxyl group of one residue and the α-amino group of another to form a peptide bond. Chemically, a polypeptide chain consists of alternating α-carbon atoms and peptide groups, with amino acid side chains projecting out from each α-carbon atom. The peptide group is planar and rigid but there is free rotation about the bonds joining the peptide group to the α-carbon atoms.

If a polypeptide chain were fully extended, each amino acid residue would occupy 3.61 angstroms of length. However, a chain is rarely, if ever, fully extended, and is instead folded considerably. If no intramolecular interactions are present, a polypeptide chain would assume the random coil conformation in which angles between residues are nearly random, limited only by the requirement that the residues cannot occupy the same space. Except at high temperatures and in the presence of powerful denaturants, the random coil conformation probably does not exist. Numerous intramolecular interactions influence the conformation of a polypeptide chain. These interactions are of three main types—between peptide groups, between side chains, and between side chains and solvent. In the absence of any interactions involving the side chains, hydrogen bonds form between the carbonyl group of a peptide group and the amide hydrogen of a peptide group three amino acid units along the chain; that is, each peptide group is hydrogen-bonded to two other peptide groups, one three units ahead and the other three units behind. This hydrogen-bonding causes the polypeptide chain to form a right-handed helix called the α helix. Two segments of a chain can also align side by side in extended form with hydrogen bonds between peptide groups on the two parallel segments. Such a conformation, which is quite rigid, is called a β structure. If the two segments have the same amino-to-carboxyl polarity, the β structure is

parallel; if the polarities are opposite, the β structure is antiparallel. Most proteins contain both α-helical and β-structural segments. If most of the amino acid residues are in either the α-helical or β conformation, the protein will be long and thin and is said to be a fibrous protein. Hair and silk are examples of fibrous proteins. If the segments are short and interspersed with either randomly coiled segments or regions in which interactions between side chains predominate, the chain is compact and is said to be globular. Most proteins, and all enzymes, are globular.

Numerous interactions involving the side chains cause folding. These are: ionic bonds between charged groups in acidic and basic amino acids; hydrogen bonds between hydroxyl, carbonyl, and amino groups; hydrophobic clustering between hydrocarbon side chains of isoleucine, leucine, phenylalanine, and valine; base stacking between phenylalanines; metal-ion coordination complexes between amino, hydroxyl, carbonyl, ring nitrogens, and pairs of SH groups. Covalent disulfide bonds between the SH groups of cysteine are often present and sometimes are the major determinant of the folding of a particular chain. Clusters of side chains often participate in van der Waals bonds. Amino acids in α-helical regions are always nearby in the chain. However, β interactions and most side chain interactions involve residues that are distant along the chain.

A general pattern of folding is that hydrophobic amino acids tend to be within the molecule, isolated from solvent, and polar amino acids tend to be on the surface. Hydrophobic patches on the surface often cause folded chains to aggregate, forming multisubunit proteins. Other interactions between groups on the surface can also cause intermolecular aggregation. Regions on the surface or in clefts in the surface often consist of clusters of amino acids that can bind small molecules. Usually the binding site is either a combination of charged groups or a surface whose shape is complementary to the surface of the small molecule, thereby allowing an attractive van der Waals interaction to occur, or a combination of the two. Often the small molecule does not match perfectly, and in an initial stage of weak binding the protein changes shape slightly to accomodate it; this phenomenon is called induced fit. If matching is initially perfect, the phenomenon is called steric fit. A protein may have multiple binding sites for the same or different molecules, and occupation of one site can affect the affinity of other sites for potential molecules that bind to the site. The affinity can increase or decrease, depending on the particular protein and on the molecule binding. This phenomenon is called allostery. Binding sites may have catalytic activity, as in the enzymes.

Enzymes are proteins with catalytic activity. They are highly specific with respect to the reaction catalyzed and typically accelerate the reaction by a factor of 10^8 to 10^{10}. An important step in catalysis is the formation of an enzyme-substrate complex, in which the substrate molecule binds to the active site of the enzyme. The strength of the binding is measured by the Michaelis constant, the substrate concentration at which the reaction velocity is half maximal. In the lock-and-key mode of binding the substrate fits in the binding site without any change in shape of the site; more common is the induced-fit mode in which the binding site and the protein itself is altered slightly by the binding. Catalysis results from numerous mechanisms, not described in this book.

ATCase is an allosteric enzyme involved in a pathway in which glutamine, carbon dioxide, ATP, and aspartate ultimately form CTP. The step catalyzed is a reaction between carbamoyl phosphate and aspartate to form carbamoyl aspartate. ATCase is a multisubunit enzyme possessing binding sites for ATP, aspartate, and CTP. ATP binds to the protein and causes a conformational change in some of the subunits that leads to activation of the enzyme. CTP binds to other sites and causes shape changes that inactivate the protein.

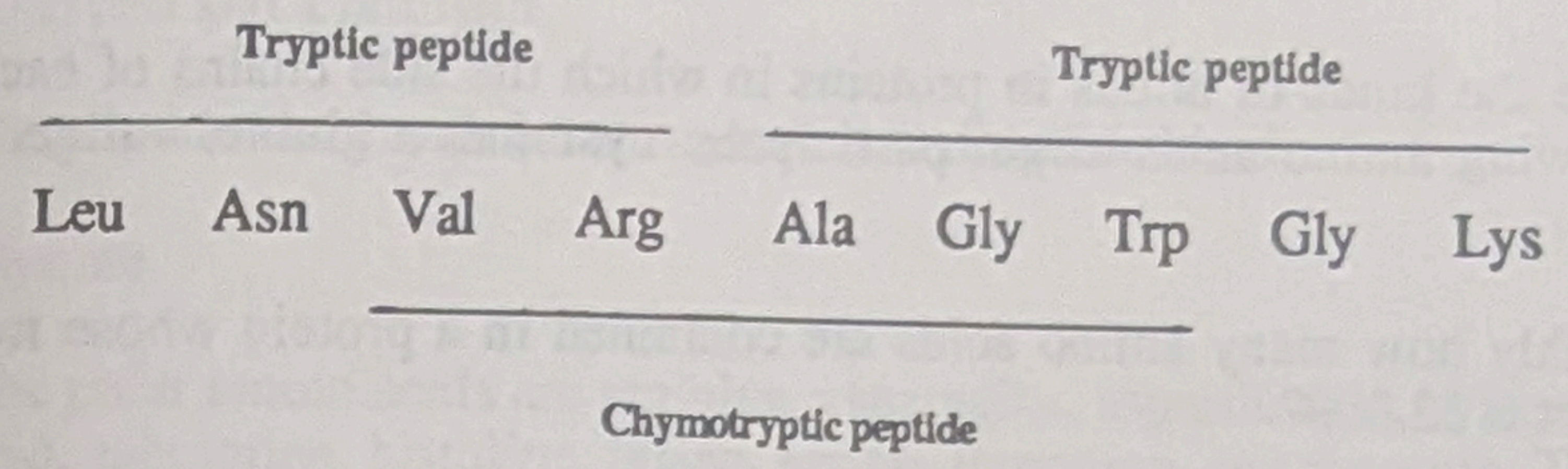
Immunoglobulin G (IgG) is a multisubunit protein that comprises one of the classes of antibodies of the immune system. It has four subunits: two identical L chains and two identical H chains. Each L chain is bound to one H chain by disulfide bonds, and the two H chains are similarly bound. The amino ends of a pair of interacting L and H chains form a binding site for the antigen. Thus, each IgG molecule has two identical antigen-binding sites. Each L chain has an amino acid sequence in the carboxyl half of the molecule that is the same for all IgG molecules; this is called the constant or C region. The sequence of the amino half of the chain varies from one antibody to the next and is called the variable or V region. H chains also have a constant and a variable region. The antigen binding site is formed from the variable regions. The plethora of variable regions accounts for the enormous number of antibody types.

BOLD TERMS

α helix, active site, affinity, allosteric proteins, antibodies, antigen-antibody reaction, antigens, antiparallel, β-pleated sheet, β structure, binding site, constant (C) region, disulfide, domains, end-product inhibition, enzyme-substrate, ES complex, fibrous protein, globular proteins, H chain, IgG, immunoglobulin fold, induced fit, L chain, lock-and-key, Michaelis constant, parallel,proteases, proteolytic enzymes, renaturation, secondary structure, steric fit, substrate, subunits, tertiary folding, tertiary structure, turnover number, V_M, variable (V) region

ADDITIONAL INFORMATION

Determination of the amino acid sequence is essential in understanding the detailed structure of a protein. This is accomplished easily by breaking the molecule into short fragments having at most fifteen amino acids, purifying each fragment, and determining the sequence of each one. The sequence of a fragment is obtained by a procedure known as the Edman degradation. Phenyl isothiocyanate reacts exclusively with the N terminus of a polypeptide. By appropriate choice of hydrolytic conditions the labeled N-terminal peptide can be removed and later identified. The next N-terminal amino acid is then reacted, removed, and identified. This process is repeated until the sequence of the peptide is obtained and then it is carried out for each peptide. The problem is then to determine the order of the peptides. This is done by cleaving the protein in several ways so that there is more than one set of polypeptides. For example, suppose two polypeptides obtained by hydrolysis with the proteolytic enzyme trypsin have the sequences Ala-Gly-Trp-Gly-Lys- and Leu-Asn-Val-Arg, and a fragment obtained by hydrolysis with chymotrypsin is Val-Arg-Ala-Gly-Trp; these sequences can be matched as follows:



Usually, by obtaining three sets of peptides and sequencing each set, the puzzle can be solved. For determining the position of disulfide bonds and the sequences for proteins consisting of two polypeptide chains held together by disulfide bonds (for example, insulin), more complicated procedures are required.

Denaturation of proteins often yields information about protein structure. Several procedures are available, three examples of which are described:
Example 1: Iodination of tyrosine. If a tyrosine is on the surface of a protein, it can react with iodine; internal tyrosines are nonreactive. Consider a protein for which chemical analysis indicates that it contains six tyrosines. A solution of the protein is divided into two parts; one is denatured and the other is untouched. Each sample is treated with radioactive iodine. If the radioactivity of the iodinated native sample is one-third of that of the sample iodinated after denaturation, we may conclude that $(1/3)6 = 2$ tyrosines are on the protein surface and $6 - 2 = 4$ are internal.
Example 2: Molecular weight determination. If the addition of a denaturant causes an *n*-fold reduction in molecular weight and a single species is present, the protein must contain *n* subunits.
Example 3: Sedimentation. The sedimentation coefficient of a macromolecule is determined in part by the shape of the molecule, as explained in Chapter 3. Since a denatured protein can be considered to be a random coil, measurement of *s* before and after denaturation can indicate whether the native form is more or less compact than a random coil. For example, if denaturation were to cause an increase in the *s* of a particular protein without a change in molecular weight, the native protein is likely to be rodlike and to have an extensive α-helical region or β structure.

DRILL QUESTIONS

1 Which atoms are covalently linked in a peptide bond?

2 When a protein is hydrolyzed completely to determine the amino acid composition, two classes of covalent bonds are usually broken. What are they?

3 Which amino acid has only a hydrogen atom as its side chain?

4 Which amino acids are polar and which are nonpolar?

5 Is leucine more or less nonpolar than alanine? Why?

6 Name the kinds of bonds in proteins in which the side chains of each of the following amino acids might participate: cysteine, arginine, valine, aspartic acid.

7 Roughly how many amino acids are contained in a protein whose molecular weight is 32,000?

8 Which amino acids, polar or nonpolar, will tend to be internal and which will tend to be on the surface of the protein? Which show little preference?

9 Which amino acid, serine or alanine, will tend to be internal and which will be external?

10 About which bond in the polypeptide backbone is there no free rotation?

11 Which amino acid cannot form a typical peptide bond?

12 What chemical groups are found at the ends of a polypeptide chain?

13 Which amino acids are linked in a disulfide bond?

14 Which of the following sets of three amino acids are probably clustered within a protein? (1) Asn, Gly, Lys; (2) Met, Asp, His; (3) Phe, Val, Ile; (4) Tyr, Ser, Lys; (5) Ala, Arg, Pro.

15 In an α helix, what fraction of the peptide bonds are engaged in hydrogen bonds?

16 Which of the following bonds or interactions are essential for forming an α helix and a β structure: hydrogen bonds, van der Waals forces, hydrophobic interactions, ionic bonds?

17 Which of the following functional classes of proteins is more likely to be fibrous rather than globular: enzymes, regulatory proteins, structural proteins?

18 What type of interaction is primarily responsible for the joining of subunits?

19 Is an enzyme likely to have a very rigid configuration? Explain.

20 What is meant by induced fit, when considering enzyme activity?

21 What is the length of an α-helical polypeptide consisting of 500 amino acids?

ANSWERS TO DRILL QUESTIONS

1 A carbon and a nitrogen.

2 The peptide bond and, in most proteins, a disulfide bond.

3 Glycine.

4 The polar amino acids are arginine, asparagine, aspartic acid, cysteine, glutamic acid, glutamine, histidine, lysine, serine, threonine, and tyrosine. The nonpolar amino acids are alanine, glycine, isoleucine, leucine, methionine, phenylalanine, proline, tryptophan, and valine.

5 Isoleucine is more nonpolar than alanine, because it has a long, nonpolar side chain.

6 Cysteine—disulfide; arginine—ionic; valine—hydrophobic; aspartic acid—ionic and hydrogen bonds. All of course could be in van der Waals bonds.

7 The average molecular weight of an amino acid residue is 110. Thus, there will be 32,000/110 = 299 amino acids.

8 Those with nonpolar side chains tend to be internal. Those with polar side chains tend to be on the surface. There is little preference for glycine, which has no side chain, for the rings with a very weak charge (e.g., tryptophan) and for cysteine.

9 Alanine, which is nonpolar, will tend to be internal; serine is polar and will tend to be on the surface.

10 The peptide bond.

11 Proline is unable to form a proper peptide because it lacks a free amino group. It is, properly speaking, an imino acid.

12 An amino group and a carboxyl group.

13 Two cysteines. Methionine, which also contains an SH group, does not form disulfide bonds.

14 Set 3, in a hydrophobic cluster.

15 All of the peptide bonds.

16 Only hydrogen bonds.

17 Structural proteins.

18 The hydrophobic interaction. Van der Waals bonds can also be important.

19 No. In general, enzymes must be slightly flexible in order to adapt their shape to the substrate and to carry out the required chemical reaction.

20 The flexing of an enzyme about a substrate.

21 About 75 nm.

PROBLEMS

1 What might you guess about the solubility of a 50-unit polypeptide containing only polyphenylalanine? Only tyrosine?

2 Most proteins completely unfold when organic solvents are added. Explain this phenomenon.

3 How would you expect the ratio of polar to nonpolar amino acids to change, on the average, with increasing size of a typical protein? (Think about the relative change in surface area and volume.)

4 Glycine is often found in regions of a polypeptide that are very tightly folded. Why might this be the case?

5 Which of the molecules, polyalanine and polyaspartic acid, will have its three-dimensional shape altered by a change in pH, and what changes will occur?

6 Some of the polar amino acids are often found inside the protein but generally in clusters. Will cysteine also show a preference? How does cysteine differ from all other amino acids with respect to its interaction with another amino acid?

7 Is polyalanine more extended in 0.3 M NaCl than in distilled water? Why?

8 Is polyglutamic acid more extended in 0.3 M NaCl than in distilled water? Explain.

9 The shape of a protein changes dramatically when the NaCl concentration is changed from 0.05 M to 0.3 M. What does this suggest about the types of interactions that determine the folding?

10 Do you think that all naturally occurring proteins probably contain at least one residue of each amino acid?

11 What kinds of three-dimensional configurations might the following peptides have: (1) Gly-Asp-Met-Ala-Ala-Glu-Val; (2) Gly-Phe-Ile-Gly-Asp-Phe-Gly; (3) Pentaphenylalanine; (4) decaglutamic acid? How would you expect any of these structures to vary with pH?

12 What would you guess to be the environment of a glutamine that is internal? What is the environment of an internal lysine?

13 The amino and carboxyl groups of most amino acids are hydrogen-bonded to water. In a dipeptide the peptide group is also hydrogen-bonded to water. For which of the following structures are the peptide groups not bound to water: α helix, β structure, random coil?

14 What is the structure of polyarginine at pH 7? How might the structure be changed?

15 A protein containing 12 amino acids is cleaved by trypsin. The peptides Met-Trp-Pro-Arg, Val-Ala-Phe-Leu-Lys, and Val-Met-Gly are produced. Treatment of the original protein with carboxypeptidase (an exopeptidase) yields mostly Gly and some Met. A protease that cleaves only at the amino-terminal side of Ala yields a pentapeptide and a heptapeptide. What is the amino acid sequence of the protein?

16 State whether the following polypeptides would aggregate to form a multisubunit protein consisting of identical subunits and give the reason for your conclusion. (1) The folded molecule contains two distinct surface regions whose shapes are complementary and no polar amino acids are nearby. (2) There is a large hydrophobic cluster that is in a crevice just below the surface. (3) A large hydrophobic patch is flanked by two lysines. (4) The surface has a region in which positively and negatively charged amino acids alternate and are in a linear array.

17 Which of the following are true statements: (1) enzymes affect the direction of a chemical reaction; (2) enzymes alter the speed of a chemical reaction; (3) enzymes are rarely, if ever, consumed in chemical reactions; (4) a particular enzyme can catalyze reactions involving different substances; (5) a particular enzyme frequently carries out different types of reactions involving a single chemical group?

18 Why do most enzymes lose activity if dissolved in distilled water?

19 What kinds of forces might hold together a protein containing identical subunits?

20 A particular enzyme has a molecular weight of 60,000. When dissolved in 7 M urea, the molecular weight drops to 20,000. When the urea is removed, the molecular weight is again 60,000 and the enzymatic activity is the same as before urea treatment. A mutant enzyme shows the same changes in molecular weight but there is no enzymatic activity either before or after the urea cycle. If equal weights of the normal and mutant enzymes are mixed and treated with urea, and then the urea is removed, what can you say about the requirements for enzymatic activity if the activity is now 50% of the starting value? 87.5%? 12.5%?

21 An enzyme has been dissociated into four identical subunits. You want to test for the enzymatic activity of the individual subunits and thus must be sure that there are no tetramers remaining in the sample. (a) What technique, described in Chapter 3, would you choose to free the monomers from the tetramers? (b) How would you know where the tetramer would be, if it were present?

22 Many enzymes are aggregates of identical subunits. Explain the biological advantages of this phenomenon.

23 The activity of many enzymes is strongly dependent on pH and is maximal at a particular pH. Furthermore, the range of pH values at which there is significant activity is often only two to three pH units. Frequently there is no detectable conformational change in this pH range, so it is unlikely that a gross conformational change accounts for this behavior. (a) Suggest an explanation for the dependence on pH. (b) What might you conclude about the active site of an enzyme if the activity increases from pH 5 to pH 7 and then remains constant (that is, it has no maximum) until pH 11, at which value the activity abruptly drops to zero?

ANSWERS TO PROBLEMS

1 Since polyphenylalanine contains no polar groups except at the termini, it should be quite insoluble. In contrast, the hydroxyl group of the ring of tyrosine is polar, so polytyrosine should be charged, and hence soluble.

2 In organic solvents the hydrophobic interaction, which is very important in protein folding, breaks down almost completely.

3 As the size of an object increases, the ratio of its surface area to its volume decreases. Since polar amino acids are more often located on the surface than within a folded polypeptide chain and the reverse is true of nonpolar amino acids, the ratio of polar to nonpolar amino acids also decreases.

4 The side chain of glycine (a hydrogen atom) is so small that there is no steric interference with tight folding.

5 Polyalanine is uncharged so it will stay in the α-helical configuration. Polyaspartic acid will be an α helix only below pH 3.8 when the carboxyl group of the side chain is not ionized.

6 Two cysteines can form a covalent bond called a disulfide bond; when this bond forms, the two cysteines are internal.

7 Alanine is nonpolar and the side chain will tend to form a hydrophobic cluster. This will be enhanced in the more charged 0.3 M NaCl solution.

8 Glutamic acid is charged, so the residues will repel one another unless there are salts to neutralize the charges. Thus, it will be more extended in distilled water.

9 A major stabilizing force must be an ionic bond, since these bonds are disrupted when the salt concentration is added.

10 Generally speaking, if the number of amino acids is greater than about 100, all would tend to be present, just from statistical considerations. For small proteins several amino acids might be missing. However, even in the large proteins, certain amino acids, such as cysteine, might be absent. Proteins that form large insoluble structures, such as tendon, skin, and virus coat proteins, frequently lack certain amino acids.

11 Think about the amino acids that would tend to form ionic bonds and hydrophobic bonds. The ionic bonds in (1) and (2) will be pH-sensitive. The hydrophobic bonds will not. (2) will have an internal hydrophobic cluster of isoleucine and two phenylalanines. (3) will stack with the rings parallel. (4) will be extended when the side chains are charged and an α helix when they are not charged; thus, this structure will also be pH-sensitive.

12 Both will tend to be near an amino acid of opposite charge.

13 The α helix and β structure, because in these structures the peptide groups are hydrogen-bonded to one another.

14 An extended coil but not an α helix, because arginine is charged. In strong alkali (pH > 13) it is uncharged and the α helix should predominate.

15 Val Ala Phe Leu Lys Met Trp Pro Arg Val Met Gly.

16 (1) Yes. A van der Waals attraction would favor aggregation. This would be aided by a hydrophobic interaction if the side chains on the surface were very nonpolar. (2) No. For geometric reasons the hydrophobic regions probably cannot come into contact. (3) No. Charge repulsion will effectively counteract any tendency to form a hydrophobic cluster. (4) Yes. If the geometry is appropriate, the alternation of unlike charges will allow unlike charges in the two

molecules to attract.

17 Statements 2, 3, 4, and 5. Note that statement 4 is true only if the substances have a common chemical feature.

18 Charged groups, normally neutralized by ions present in the solvent, interact and thereby change the shape of the protein from an active to an inactive form.

19 Hydrophobic interactions are the most common, but all types of bonds can do it.

20 The enzyme has three subunits if, in urea, the molecular weight drops threefold. When active and inactive forms are mixed, dissociated, and reassociated, the subunits will reassociate at random to produce molecules having some active and some inactive subunits; those having 0, 1, 2, and 3 active subunits arise at frequencies 1/8, 3/8, 3/8, and 1/8, respectively. If the activities are 12.5%, 50%, and 87.5%, there must be 3, 2, and 1 active subunits, respectively, if the molecule is to be active.

21 (a) Gel electrophoresis. Gel chromatography would also work. (b) Add proteins of known molecular weight and obtain a standard curve in which either the distance migrated (for electrophoresis) or the volume of the effluent (for chromatography) is plotted against molecular weight.

22 Activity can be regulated by dissociation and association of subunits. For example, in a protein solution at low concentration, an inactive monomer will be the major species. Also, less genetic information is required to produce the complete protein.

23 (a) The charge on one or more amino acids in the active site is changed. (b) At pH 11 the protein denatures. There is an amino acid in the active site whose charge changes between pH 5 and 7.

MORE CHALLENGING QUESTIONS

1 Most enzymes lose all activity when treated with formaldehyde. Why? Hint: Think about the chemical reactivity of formaldehyde.

2 The two least frequent amino acids in protein X are tryptophan and methionine. Protein X contains 0.50 mole % tryptophan and 1.50 mole % methionine. (a) What is the minimum number of amino acids per molecule of protein X? (b) If protein X has a sedimentation coefficient that is slightly less than that of a protein having 267 amino acids, what is the probable number of methionines per molecule of protein X?

3 The enzyme carboxypeptidase can remove the terminal amino acid of a protein by cleaving the peptide bond between the terminal amino acid and the penultimate amino acid. After the terminal amino acid is removed, the penultimate one can be removed. A protein is treated with carboxypeptidase for a limited period of time and only alanine is recovered, in the ratio of two alanines per protein molecules. What is the amino acid sequence of the terminal tripeptide?

4 A protein is treated with carboxypeptidase and then precipitated from solution. The supernatant is analyzed and found to consist of 82 mole % valine and 18 mole % alanine. By a variety of criteria you know that at least 99.9% of the original sample consists of a single protein. How do you explain the production of two amino acids by carboxypeptidase?

5 Most enzymes lose activity if a dilute solution is shaken so violently that foaming occurs. This can happen even when there is no decrease in molecular weight. Why is the activity lost?

6 In general, a diploid cell containing the + and – alleles encoding an enzyme E has an E^+ phenotype. You have isolated a particular mutant that yields an E^- phenotype even when the + allele is present. Suggest a possible explanation for the observed phenotype.

7 No amino acid side chain is a good electron acceptor, yet many enzymes catalyze reactions in which the enzyme serves as an electron acceptor. How is this accomplished?

8 An enzyme is known to require a high concentration of the magnesium ion for activity. If the magnesium ion is removed, the protein is irreversibly denatured. In attempting to develop a purification scheme for the enzyme, you try both ion-exchange and gel chromatography. In both cases the enzyme loses activity. Explain why that might happen. In view of your explanation, what modifications could you make to improve the situation?

9 Why do enzymes vary so much in size? Would you expect the enzymes involved in glucose metabolism and in amino acid metabolism to be relatively large or relatively small?

ANSWERS TO MORE CHALLENGING PROBLEMS

1 Formaldehyde reacts with amino groups, thereby preventing hydrogen-bonding.

2 (a) Since each molecule of protein X must contain at least one residue of the rarest amino acid, the number of amino acids in X cannot be less than 100/0.5 = 200. (b) The value of *s* is an approximate measure of the molecular weight, which, in turn, is an approximate measure of the total number of amino acids in

a protein. An *s*-value slightly less than that of the particular protein implies a number of amino acids less than 267. Evidently, protein X does contain 200 amino acids rather than some integral multiple of 200, and the number of methionines per molecule must be simply 3.

3 Once the enzyme has cleaved a terminal amino acid, it may then either cleave the next amino acid or remove the terminal amino acid from a second protein. Hence, when two amino acids per protein, on the average, have been removed, there are proteins which have had one, two, and three amino acids removed. Since only alanine has been found, the terminal tripeptide must consist of three alanines.

4 The C-terminal amino acid is valine. Once this is removed, the enzyme can remove the adjacent amino acid.

5 Foaming increases the surface-to-volume ratio of the solution. This increases the probability that an enzyme molecule will be at the surface and subjected to disruption by the force of surface tension. It is worthwhile to determine for yourself how these forces can destroy activity. If the enzyme contains essential sulfhydryl groups in or near the active site, inactivation may result from the greater rate of oxidation of these groups in the highly aerated foam.

6 The enzyme could be a multisubunit protein in which one defective subunit is sufficient to eliminate enzymatic activity.

7 Metal ions or small molecules are part of the active enzyme; usually, these are the electron acceptors.

8 If the binding of the ion is not too tight, a high concentration of the ion is needed to have bound ions. The ion exchanger will remove these ions. In gel chromatography the ions move more slowly than the protein and will be separated. The solution to the problem is to try gel chromatography with a column equilibrated with the ion and to elute with a solution of the same ion concentration.

9 Enzymes whose activity is controlled by the presence of small molecules (for example, allosteric enzymes) must contain several binding sites and therefore tend to be large. A typical protein with a single active site usually has a molecular weight of 20,000 to 60,000. The enzymes participating in glucose metabolism are active most of the time and hence do not need regulatory binding sites; hence they are, on the average, fairly small. The enzymes involved in amino acid metabolism are not needed if the amino acids are supplied from outside the cell. Thus, they are regulated; the particular enzymes in each pathway that are regulated tend to be larger than the others.

CHAPTER 7

MACROMOLECULAR INTERACTIONS AND THE STRUCTURE OF COMPLEX AGGREGATES

CHAPTER SUMMARY

Many macromolecules, such as the majority of the enzymes, exist as free intracellular entities. However, numerous macromolecules are in the form of aggregates. The components of these aggregates may be of the same type, as in the complex polysaccharide structure of the bacterial cell wall and in multisubunit protein assemblies, or can be of different classes, as in nucleoproteins. Such aggregates can be intracellular (chromatin, muscle sarcomeres, mitotic spindle), can surround the cell (cell membranes, cell walls), or extracellular (tendon, bone, hair, insect shells).

One of the most abundant protein structures in nature is collagen, which is found in skin, tendon, the dura mater of the brain, and bone. Collagen has an unusual amino acid composition, consisting primarily of glycine and proline. Many of the prolines are hydroxylated. Single polypeptide chains (procollagen) interact to form a triple-stranded helix, called tropocollagen. The surface of tropocollagen contains numerous positively and negatively charged amino acids which engage in ionic bonds, producing side-to-side aggregation of the helices. The pattern of these charged amino acids causes the tropocollagen helices to aggregate such that the end of one helix is one-fourth the distance along the adjacent chain. This property enables the tropocollagen helices to form long collagen fibers without the necessity of end-to-end interaction. Cross links between hydroxyprolines in tropocollagen helices, plus the quarter stagger, results in the formation of long fibers with great tensile strength.

Another well-characterized structure is that of the *E. coli* chromosome. In this structure a long supercoiled DNA molecule is folded by virtue of binding to specific proteins to form a multiply looped fairly symmetric structure, whose cross-sectional dimension is a few micrometers. This small volume contrasts with the dimensions of the extended nonsupercoiled circle, which has a diameter of roughly 350 micrometers. The individual domains are physically isolated from one another in the sense that a single-strand break in the DNA causes a loss of supercoiling of only a single loop rather than of the entire DNA molecule.

a protein. An *s*-value slightly less than that of the particular protein implies a number of amino acids less than 267. Evidently, protein X does contain 200 amino acids rather than some integral multiple of 200, and the number of methionines per molecule must be simply 3.

3 Once the enzyme has cleaved a terminal amino acid, it may then either cleave the next amino acid or remove the terminal amino acid from a second protein. Hence, when two amino acids per protein, on the average, have been removed, there are proteins which have had one, two, and three amino acids removed. Since only alanine has been found, the terminal tripeptide must consist of three alanines.

4 The C-terminal amino acid is valine. Once this is removed, the enzyme can remove the adjacent amino acid.

5 Foaming increases the surface-to-volume ratio of the solution. This increases the probability that an enzyme molecule will be at the surface and subjected to disruption by the force of surface tension. It is worthwhile to determine for yourself how these forces can destroy activity. If the enzyme contains essential sulfhydryl groups in or near the active site, inactivation may result from the greater rate of oxidation of these groups in the highly aerated foam.

6 The enzyme could be a multisubunit protein in which one defective subunit is sufficient to eliminate enzymatic activity.

7 Metal ions or small molecules are part of the active enzyme; usually, these are the electron acceptors.

8 If the binding of the ion is not too tight, a high concentration of the ion is needed to have bound ions. The ion exchanger will remove these ions. In gel chromatography the ions move more slowly than the protein and will be separated. The solution to the problem is to try gel chromatography with a column equilibrated with the ion and to elute with a solution of the same ion concentration.

9 Enzymes whose activity is controlled by the presence of small molecules (for example, allosteric enzymes) must contain several binding sites and therefore tend to be large. A typical protein with a single active site usually has a molecular weight of 20,000 to 60,000. The enzymes participating in glucose metabolism are active most of the time and hence do not need regulatory binding sites; hence they are, on the average, fairly small. The enzymes involved in amino acid metabolism are not needed if the amino acids are supplied from outside the cell. Thus, they are regulated; the particular enzymes in each pathway that are regulated tend to be larger than the others.

CHAPTER 7

MACROMOLECULAR INTERACTIONS AND THE STRUCTURE OF COMPLEX AGGREGATES

CHAPTER SUMMARY

Many macromolecules, such as the majority of the enzymes, exist as free intracellular entities. However, numerous macromolecules are in the form of aggregates. The components of these aggregates may be of the same type, as in the complex polysaccharide structure of the bacterial cell wall and in multisubunit protein assemblies, or can be of different classes, as in nucleoproteins. Such aggregates can be intracellular (chromatin, muscle sarcomeres, mitotic spindle), can surround the cell (cell membranes, cell walls), or extracellular (tendon, bone, hair, insect shells).

One of the most abundant protein structures in nature is collagen, which is found in skin, tendon, the dura mater of the brain, and bone. Collagen has an unusual amino acid composition, consisting primarily of glycine and proline. Many of the prolines are hydroxylated. Single polypeptide chains (procollagen) interact to form a triple-stranded helix, called tropocollagen. The surface of tropocollagen contains numerous positively and negatively charged amino acids which engage in ionic bonds, producing side-to-side aggregation of the helices. The pattern of these charged amino acids causes the tropocollagen helices to aggregate such that the end of one helix is one-fourth the distance along the adjacent chain. This property enables the tropocollagen helices to form long collagen fibers without the necessity of end-to-end interaction. Cross links between hydroxyprolines in tropocollagen helices, plus the quarter stagger, results in the formation of long fibers with great tensile strength.

Another well-characterized structure is that of the *E. coli* chromosome. In this structure a long supercoiled DNA molecule is folded by virtue of binding to specific proteins to form a multiply looped fairly symmetric structure, whose cross-sectional dimension is a few micrometers. This small volume contrasts with the dimensions of the extended nonsupercoiled circle, which has a diameter of roughly 350 micrometers. The individual domains are physically isolated from one another in the sense that a single-strand break in the DNA causes a loss of supercoiling of only a single loop rather than of the entire DNA molecule.

The supercoiled state is maintained by a balance between two topoisomerases. RNA is often found in the condensed chromosome, but this is not thought to be an essential part of the structure.

In eukaryotes DNA is bound to histones to form chromatin. Four histones, H2A, H2B, H3, and H4, form an octameric disc around which DNA makes 1-3/4 turns, forming a unit called a nucleosome. The entire DNA molecule is wrapped around numerous octameric discs, so the DNA may be likened to a string of nucleosome beads. The nucleosomes are not in contact with one another but are separated by a segment of DNA termed a linker. A fifth histone, H1, is located in part of the linker region and may be responsible for bringing the nucleosomes nearer. The chromatin (that is, the string of beads) is folded in a series of stages. The first stage is the formation of a helical solenoid called the 30-nm fiber. This fiber is then folded further but the manner of folding is not yet known. Each stage of folding condenses the DNA until a chromosome is formed. Further condensation of an unknown kind occurs during mitosis and meiosis, when chromsomes become visible in the light microscope. This condensation probably involves an interaction (of an unknown type) with a complex protein structure called the scaffold; RNA may also be involved in this condensation.

The centromere is a special region of each chromosome used for attachment of the chromosome to the mitotic spindle. The structure of the centromere is not well known but it is certainly somewhat different than typical chromatin with respect to histone composition and arrangement. A particular segment of the centromere, called the kinetochore, is the actual site of attachment of a fiber that is part of the spindle system. The termini of chromosomes, called telomeres, also have a special structure. The DNA consists of a series of short sequences that are repeated 20–70 times. Single-strand breaks and gaps are present in this part of the DNA, and at the end of the DNA strand the two strands are joined to form a hairpin structure. Telomeres are thought to be responsible for completion of replication of the linear chromosomal DNA molecule, but neither the mechanism by which this occurs nor the structure of the telomere are known.

Many proteins are known that bind to specific base sequences in DNA. A common feature in the binding is the insertion of α helices of the protein into the major groove of the DNA. Binding specificity is a result of interactions between particular amino acids in the α helix with particular bases in the major groove. It is also common for the protein to bind only to one side of the DNA molecule. This type of binding has been observed for the Cro, cI, and cII proteins of phage λ and related phages, for the cAMP-CRP protein of *E. coli,* and for the EcoI restriction enzyme. The Klenow fragment of *E. coli* DNA polymerase I, which does not bind to specific base sequences, also binds to DNA via α helices. Analysis of the EcoRI-DNA complex shows that the DNA helix is deformed somewhat by the binding. Although such deformation has not yet been observed in other binding systems, it is expected that it, or a variant thereof, will be observed.

Biological membranes are assemblies of lipids and proteins. The basic structure is a lipid bilayer, in which the polar regions of lipids are on the two surfaces of the bilayer and the nonpolar hydrocarbon tails are internal. The bilayer contains other molecules, which give the membrane specific properties. For example, cholesterol increases the rigidity of a membrane. Various proteins contained in the membrane are involved in the transport of small molecules across the membrane; in the absence of these proteins the membrane would normally be impermeable to these small molecules. These internal proteins, and proteins that are both internal and are exposed on the membrane surface, are called integral membrane proteins. Peripheral proteins are bound to the membrane surface, usually to the integral membrane proteins. Membranes are fairly fluid structures and both internal and surface proteins tend to drift around, possibly randomly. When two cells containing different proteins are fused, the integral membrane proteins of the new cell drift until they are distributed uniformly through the membrane. When cell membranes are disrupted, membrane fragments often close on themselves, forming vesicles. The vesicles contained the solution in which they were suspended during their formation. Vesicles formed in the presence of various molecules are useful in studying the permeability of the membrane to particular molecules.

Various mechanisms exist for the assembly of complex structures. Some mechanisms utilize enzymes and scaffolding proteins that direct aggregation and determine the sequence of events. Other structures self-assemble. The plant virus, tobacco mosaic virus, which consists of a single RNA molecule and a large number of identical protein subunits, is an example of a self-assembled structure. The process begins with aggregation of the proteins to form a two-layered disc. A hole in the center of the disc spontaneously binds to a particular base sequence in the viral RNA. The disc then undergoes a conformational change and forms a lock-washer structure. A second disc then joins the structure again by binding of RNA to the hole in the disc, and this disc assumes the lock-washer form. The process continues repeatedly until the RNA molecule is completely contained within the stack of discs.

BOLD TERMS

bilayer, centromere, chromatin, chromosome, core particle, fibroblast, fluid mosaic model, freeze-fracture electron microscopy, histones, integral membrane proteins, linker DNA, micelle, nucleoid, nucleosome, peripheral protein, procollagen, quarter-staggered array, recognition helix, scaffold, self-assembly, telomere, topoisomerase I, triple-stranded helix, tropocollagen, vesicle

ADDITIONAL INFORMATION

When *E. coli* cells are lysed in certain ways, the released DNA does not have the appearance of either extended strands or of supercoils but the strands appear wider and condensed by the binding of molecules that are presumably proteins. This suggests that *E. coli* DNA may exist in a chromatin-like form. At present there is little understanding of this phenomenon at a molecular level but several DNA-binding proteins have been isolated that are highly basic and superficially resemble eukaryotic chromosomal proteins. A suggestion that these may be structural proteins is the large number of copies per cell. For example, 40,000 copies of one of these proteins, called HU, is present in each cell. Protein HU is interesting in that it stimulates DNA replication in some *in vitro* systems. Other proteins that may be involved in condensing bacterial DNA are H (30,000 copies per cell), HLP1 (20,000 copies/cell), and P (quantity unknown but resembling in amino acid composition a protein in sperm that plays the same role as histones). Whether these proteins are involved in forming bacterial "chromatin" is unknown; needless to say, they are actively being studied.

The *E. coli* cell wall is an interesting and complex structure. It consists of three layers, in order from outside to inside, called the outer membrane, the peptidoglycan, and the plasma membrane; the latter has a structure much like the membranes described in the text of Chapter 7 and provides the major permeability barrier for the cell. The outer membrane has a very complex structure and will not be described further. Treatment of cells with various detergents alters the structure of the membranes and increases the permeability to many small molecules. The peptidoglycan layer is an unusual substance. It is a polymer consisting of both sugar and peptide units. Two sugars, *N*-acetylglucosamine and *N*-acetylmuramic acid (a sugar to which the tetrapeptide Ala-Gln-Lys-Ala is covalently linked), alternate to form a linear polysaccharide chain with a tetrapeptide side chain. A pentaglycyl peptide forms cross links between the tetrapeptides joining the lysine of one unit to the terminal alanine of another (Figure ADD-1). This cross-linking forms a large planar sheet. The remarkable feature of peptidoglycan is that the sheet ultimately closes on itself to form one enormous saclike macromolecule which encloses the entire inner membrane and cytoplasm of the bacterium.

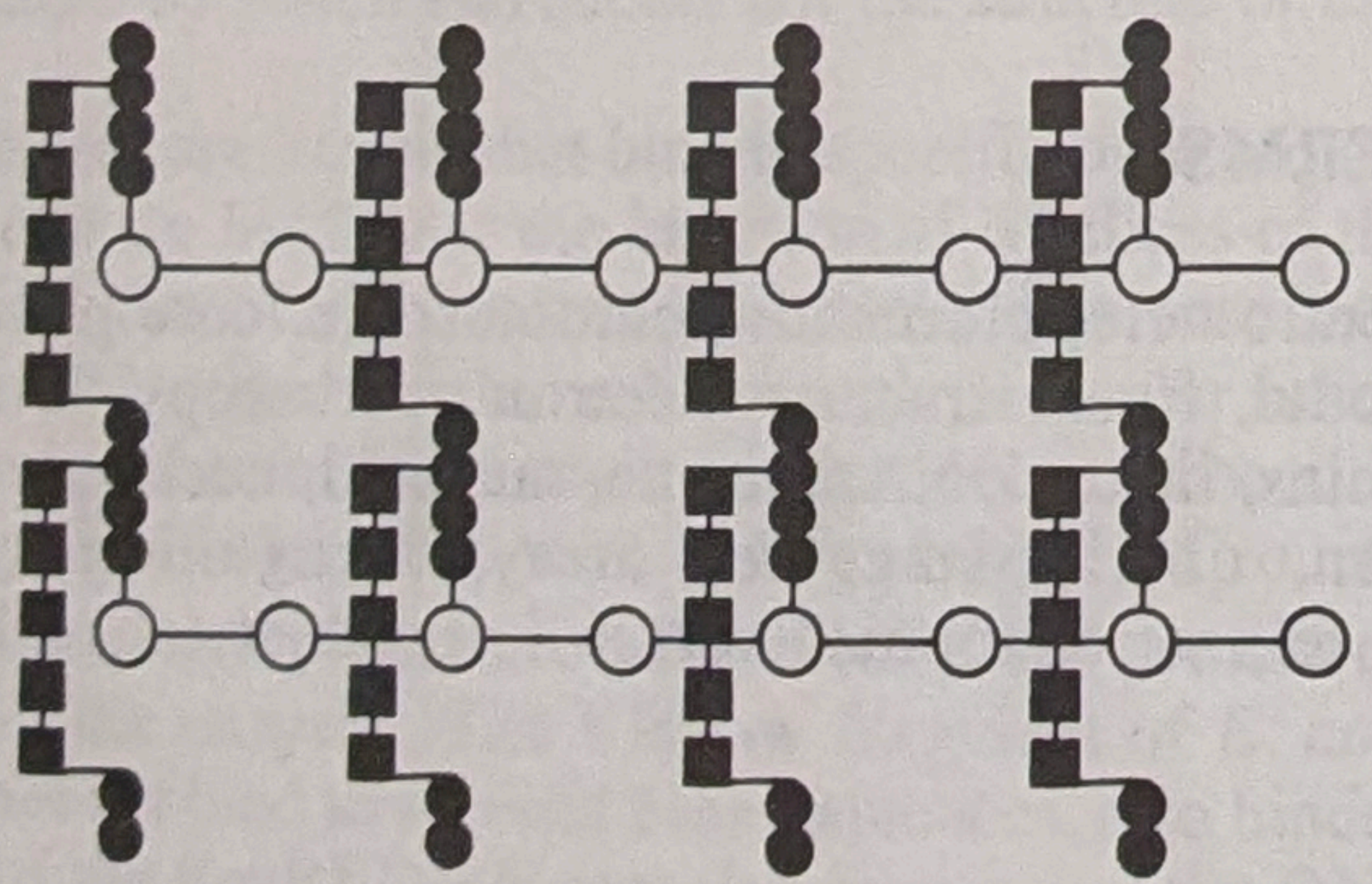


Figure ADD-1 Peptidoglycan model. Sugars—large circles; tetrapeptides—small circles; pentaglycine bridges—five linked squares

Peptidoglycan is the site of attack of the enzyme lysozyme (discussed in Chapters 5 and 6). Since peptidoglycan is the only rigid component of the cell wall, a lysozyme-treated bacterium assumes a spherical shape (whether the bacterium is initially spherical or rodlike, or has any other shape); a cell treated in this way is called a spheroplast. The antibiotic penicillin interferes with the synthesis of peptidoglycan. Thus, if a bacterium is allowed to grow in the presence of penicillin, it enlarges without growth of the strong peptidoglycan layer; the volume exceeding that which can be held by the weaker cell membranes and the cell bursts, which accounts for the lethal effect of penicillin. This bursting is the basis of the penicillin-selection technique (Chapter 11, *Study Guide*).

It is interesting to consider why bacteria evolved with a multilayered cell wall, in contrast with most animal cells, which are bounded only by a membrane similar in structure to the innermost layer of the bacterial cell wall.The most likely reason is that bacteria are in contact with many noxious environmental substances which can damage a variety of biological molecules. A multiple layer that consists of different components ensures that if one layer is damaged, the physical integrity of the cell need not be lost. Free-living eukaryotes such as the fungi also have complex multilayered cell walls. The peptidoglycan layer has also evolved with a large number of chemically protected groups; for example, some of the amino acids in the tetrapeptide are D-amino acids, which are not found in proteins. Presumably this renders the cell wall somewhat resistant to digestive enzymes in the environment.

DRILL QUESTIONS

1 How many subunits are contained in tropocollagen?

2 What amino acids are found in collagen and not in a typical protein?

3 What is unusual about the amino acid composition of collagen?

4 What is unusual about the amino acid sequence of collagen?

5 What properties are conferred on collagen by the large amount of proline?

6 What role does collagen play in formation of bone?

7 What molecules comprise the *E. coli* chromosome?

8 What role does RNA play in the structure of a bacterial chromosome?

9 What evidence supports the notion that the loops of a bacterial chromosome are in some way isolated from one another?

10 How many different types of proteins are found in a nucleosome, and what are

their names?

11 Which histone is found in chromatin but not in the nucleosome unit?

12 What is the evidence that DNA and histones are held together by ionic bonds in a nucleosome unit?

13 How many protein molecules are contained in a nucleosome core particle?

14 Which type of histone does not exist in chromatin in pairs?

15 Roughly how many base pairs of DNA are contained in a histone core particle?

16 Are the ratios of the different histones to one another the same in all cells of a eukaryotic organism? In all eukaryotic organisms?

17 Is the histone:DNA ratio the same for all cells of a eukaryotic organism? In all eukaryotic organisms?

18 Which histone is bound only to linker DNA?

19 Are histones found in all cells?

20 What would happen to a chromosome if it were put in concentrated NaCl?

21 What is the relation between the 110-A fiber and the 30-nm fiber in chromatin?

22 What kinds of genes are found in heterochromatin?

23 What is the location of a telomere on a chromosome?

24 Is there an analogue of a telomere in the *E. coli* chromosome?

25 In which groove of DNA do most DNA-binding proteins bind?

26 What feature seems to be common to most DNA-binding proteins?

27 What kinds of bonds are probably the major feature of a protein binding to a specific base sequence in DNA?

28 What bonds are probably most important in providing binding strength in DNA-protein binding? What is the evidence for this?

29 In DNA-protein binding what is meant by a kink?

30 What classes of molecules are the major components of cellular membranes?

31 Are the outer surfaces of a lipid bilayer hydrophilic or hydrophobic?

32 What structural feature must a protein have in order to be an integral membrane protein?

33 Are integral membrane proteins more or less hydrophobic than peripheral proteins?

34 When a vesicle forms, does the inner and outer side of the membrane change sides?

35 What type of molecules, polar or nonpolar, are unable to pass through a lipid bilayer that does not contain integral membrane proteins?

36 Do the integral membrane proteins have a constant location in a membrane?

37 What is the primary function of cholesterol in many cell membranes?

ANSWERS TO DRILL QUESTIONS

1 Three procollagen subunits.

2 Hydroxyproline and hydroxylysine.

3 One third is glycine and one fourth is proline, which is normally found in rather small amounts in most proteins.

4 On the average, every third amino acid is glycine.

5 Because proline does not form a typical peptide bond, the collagen monomers (procollagen) are quite extended. Proline, in its hydroxylated form, also engages in interstrand cross links.

6 The hydroxyapatite crystals, of which bone are composed, are first deposited in the gaps between longitudinally arranged tropocollagen units.

7 A single DNA molecule and numerous protein molecules.

8 Probably none.

9 One single-strand break causes loss of supercoiling of one loop rather than of the entire DNA molecule.

10 Four, the histones H2A, H2B, H3, and H4.

11 H1.

12 Chromatin is dissociated into DNA and proteins when it is placed in a solution having a high salt concentration.

13 8.

14 H1.

15 About 200.

16 There is always a 1:1:1:1 ratio of histones H2A, H2B, H3, and H4 in all eukaryotic cells because of the universality of the histone octamer. The ratio of H1 to the octamer varies slightly.

17 No, because the size of the linker varies from one cell type and from one organism to the next.

18 H1.

19 No, not in prokaryotes and not in sperm. DNA is bound to protamines in sperm.

20 The histones would no longer be able to bind to the DNA, so the entire chromosome would be disrupted and the DNA freed of the protein.

21 If a 110-A fiber is arranged in helical form with about six nucleosomes per turn, the result is a 30-nm fiber.

22 Generally, genes are not present in heterochromatin.

23 Telomeres are at the termini.

24 No, because the *E. coli* chromosome is circular and hence has no termini.

25 The major groove.

26 A pair of α helices oriented and spaced such that they can slip into two regions of the major groove.

27 Hydrogen bonds and van der Waals bonds.

28 Ionic bonds between positively charged groups on the protein and the negatively charged phosphates. Binding is usually inhibited if the salt concentration is high, which breaks down ionic bonds.

29 A kink is the deformation of the normal B-DNA structure in the region of contact with the protein.

30 Phospholipids and proteins.

31 Hydrophilic.

32 It must have a large hydrophobic part on its surface, which penetrates the lipid layer, and one or two polar regions, which extend out of the lipid layer.

33 They are more hydrophobic.

34 No.

35 Polar molecules.

36 No, they drift laterally.

37 It strengthens the membrane.

PROBLEMS

1 Collagen fibers extracted from calves are easily dissociated by suspension in 1 M NaCl. The fibers extracted from adult cows are resistant to 1 M NaCl and cannot be dissociated except by treatment with acid. Explain the difference.

2 Tropocollagen is a molecule with microscopic dimensions but is capable of forming tendons many centimeters long. What feature of the molecules allows the molecule to form such long fibers?

3 Would you expect the folding of chromatin in a eukaryotic cell to vary at different stages of the life cycle of a cell? Explain the reasons for your answer.

4 If you had never seen a chromosome with a microscope but had seen cells with nuclei, what information would nonetheless let you know that DNA must exist in a highly coiled form within cells?

5 The *E. coli* chromosome seems to have a central region to which all of the loops are joined. Suggest an explanation for such a single centralized binding region.

6 The folded *E. coli* chromosome has a high sedimentation coefficient and solutions containing it have a very low viscosity compared to most DNA solutions. Why is this so? What treatment would both lower the sedimentation coefficient and raise the viscosity?

7 A sample of identical DNA molecules each containing about 3000 base pairs per

molecule is mixed with histone octamers under conditions that allow formation of chromatin. The reconstituted chromatin is then treated with a nuclease and enzymatic digestion is allowed to occur. The histones are removed and the positions of the cuts in the DNA are identified by sequencing the fragments. It is found that the breaks have been made at random positions, and, as expected, at about 200-base-pair intervals. The experiment is then repeated with a single variation. A protein P, known to bind to DNA, is added to the DNA sample before addition of the histone octamers. Again, reconstituted chromatin is formed and digested with the nuclease. In this experiment it is found that the breaks are again at intervals of about 200 base pairs but they are localized at particular positions in the base sequence. At each position the site of breakage can vary only over a 2–3 base range. However, examination of the base sequences in which the breaks have occurred does not indicate that breakage occurs in a particular sequence—that is, each 2–3-base region in which cutting occurs has a different sequence. Explain the difference between the two experiments.

8 A DNA-binding protein binds tightly to double-stranded DNA and very poorly to single-stranded DNA. It can bind to all base sequences with equal affinity. In 1 M NaCl binding is poor. What is the probable binding site of the DNA?

9 A DNA-binding protein binds weakly to double-stranded DNA in 0.01 M NaCl and quite strongly in 1 M NaCl. Furthermore, it only binds to specific base sequences. What features of the protein might account for this effect of salt concentration?

10 The DNA from certain animal viruses, whose DNA is a covalently closed circle, can be isolated from infected cells as a circle of nucleosomes. This structure is called a minichromosome. When the histones are removed, the DNA assumes the form of a negative superhelix. What is the handedness of the coiling of the DNA around the histone octamer?

11 A particular protein molecule is known to be a regulator of gene activity in a bacterium. When the protein is purified, it is found that it binds to DNA exceedingly poorly. When the three-dimensional structure of the protein is worked out, it is found to contain a single α helix. Why do you think the isolated protein fails to show significant binding?

12 If the λ Cro protein and λ DNA are mixed, the Cro-λ-DNA complex forms. What would you think would happen if an acridine compound or ethidium bromide (both of these are intercalating agents) were added to the DNA before Cro protein was added?

13 Most DNases make single-strand breaks in DNA. However, if a DNase is added to DNA in a solution containing 1 M NaCl, the DNA molecule is completely resistant to the enzyme. Suggest an explanation.

14 Cell membranes obtained from different types of cells and even from different species can fuse. Does this imply that all membranes are formed from identical components?

15 An antibody is prepared against a peripheral protein of a cell membrane. A fluorescent molecule is coupled to the antibody in order that the antibody can be localized on the cell. The antibody is added to cells and the cells are immediately observed by fluorescence microscopy. A faint fluorescence is seen over the entire surface. Several hours later the cells are examined again and a single brightly fluorescent spot is seen. Explain this observation. [Hint: Recall that antibodies have two antigen-binding sites.]

16 Functional tobacco mosaic virus particles all have the same length. Virus-like particles can be reconstituted—in the absence of RNA—from purified virion protein. What relation would you expect between the lengths of the virus particles and the RNA-free reconstituted particles?

ANSWERS TO PROBLEMS

1 The collagen of the young animals has fewer cross links and can be dissociated by shielding of the ionic bonds by a high salt concentration.

2 The positively and negatively charged amino acids produce the quarter-stagger array, which can become indefinitely long.

3 It seems reasonable that folding would change during the replication phase, although there is no evidence for such a change. However, there clearly is a change in the degree of condensation of the chromosome between interphase, when the chromosomes are not visible by light microscopy, and the stages of mitosis, during which the chromosomes are visible. The chromosomes are certainly more condensed when visible.

4 From the total molecular weight you would know the total length of the DNA in a nucleus. Even allowing for up to 100 distinct DNA molecules per cell, the length of each molecule would still be very great compared to the diameter of a nucleus. You would be forced to conclude that each DNA molecule must fold back on itself repeatedly.

5 It probably enables the DNA molecule, whose length is about 400 times that of a bacterium, to fit neatly inside the cell. It may also be important in guiding daughter DNA molecules to daughter cells.

6 It is very compact. Removal of the RNA and protein needed to maintain the compact structure will reduce *s* and increase the viscosity.

7 In the first experiment, nucleosomes have formed at random. In the second, P has bound to a unique sequence and nucleosomes have formed by sequential addition of histone octamers from the site of P. Thus, the linkers must be at fixed distances from the unique binding sequence. Since nuclease attack occurs only in the linker regions, the cuts are made in small regions that are highly localized.

8 The fact that binding is weak in 1 M NaCl suggests an electrical attraction is involved in the binding. The charges on the bases are sufficiently weak (and mostly engaged in hydrogen bonding) that only the highly charged phosphate groups of the sugar-phosphate backbone seem to be candidates. Weak binding to single-stranded DNA is consistent with this conclusion, because a randomly coiled single strand would not present an ordered array of charges to the protein; the rigidity of double-stranded DNA provides such order. Therefore, the protein undoubtedly binds to the phosphates.

9 The structure of the protein might vary with salt concentration and have a better fit to the DNA in 1 M NaCl. Alternatively, it may bind as a dimer and the dimer form is favored over the monomer in 1 M NaCl. More likely, the protein has a negatively charged amino acid on its surface and hence is repelled by the negatively charged phosphates in the DNA. In 1 M NaCl these charges would be shielded from one another.

10 In a negative superhelix the DNA is underwound. Thus, the coiling around the histone octamer must be left-handed. This phenomenon is in fact the cause of the supercoiling of most isolated viral DNAs and of the supercoiling of histone-depleted eukaryotic chromosomes.

11 It probably exists as a dimer, thereby providing a pair of recognition helices, and you have purified it in the monomer form. A good approach to understanding the problem would be to find conditions that lead to dimerization and then test binding to DNA.

12 These molecules are intercalating agents and would move the DNA bases apart. Since the Cro protein binds stereospecifically, it is likely that the Cro-DNA complex would not form. Also, the intercalating agent might block specific contact points.

13 The DNases probably bind to the negatively charged phosphates, which are neutralized by the Na ions in 1 M NaCl.

14 No. They fuse for three reasons: (1) they all are formed from molecules having the same basic structure, (2) the membranes themselves have the same basic structure, and (3) hydrophobic interactions are nonspecific with respect to the identity of either the hydrophobic or hydrophilic parts of the components.

15 At first the antibodies are located all over the cell surface. However, because of the fluidity of the cell membrane, the peripheral proteins drift around. When an antibody bound to one protein encounters a second protein, it binds to the protein and links the proteins together. Ultimately, all of the particular peripheral protein molecules are linked, so all antibodies and hence all fluorescence is in a single spot on the cell. This phenomenon is called capping.

16 Without the RNA to fix the length, the reconstituted particles, which have a slightly different structure, grow indefinitely long.

MORE CHALLENGING PROBLEMS

1 Vitamin C is required to maintain prolyl hydroxylase, the enzyme that forms hydroxyproline, in an active form. Vitamin C deficiency causes scurvy, whose initial symptom is the appearance of bloody spots under the skin all over the body. Suggest why the red spots appear.

2 If young rats ingest seeds of the sweet pea, *Lathyris odoratus*, they develop a condition called lathyrism in which many bodily structures lose mechanical strength; for example, the tail is attached so loosely to the body that it is easily pulled off. Lathyrism is a collagen disease. However, isolated tropocollagen (the triple helix) is normal and collagen fibers have the normal band pattern. Suggest a molecular defect.

3 Mutations have not been observed that give rise to altered histones. Why should this result be expected?

4 A DNA-protein complex has been isolated and you wish to determine whether the DNA and protein are covalently linked. (a) Which of the following experimental results would give information about covalent linkage? (1) Treatment of the complex with 2 M NaCl does not dissociate the protein from the DNA; (2) treatment of the complex with a reagent that breaks down hydrophobic interactions does not dissociate the complex; (3) digestion with a protease removes all detectable protein from the DNA. (b) Describe an experimental result which would prove that there is covalent linkage.

5 At high temperatures amines catalyze breakage of phosphodiester bonds. This phenomenon has been used to study a model system for nuclease activity by using apurinic acid as a substrate. Apurinic acid is a DNA-like molecule obtained by treatment of DNA with weak acid, which causes cleavage of the *N*-glycosidic bonds between purines and deoxyribose. Thus, apurinic acid lacks all purines, although the sugar-phosphate chain is intact. The enzyme analogue is a tripeptide. Some of the following tripeptides are very effective in hydrolyzing apurinic acid. Rank the following tripeptides in order of effectiveness and give reasons for your answer: Lys-Trp-Lys, Lys-Gly-Gly, Lys-Trp-Gly, Lys-Tyr-Lys, Lys-Gly-Gly.

ANSWERS TO MORE CHALLENGING PROBLEMS

1 Collagen is a major constituent of the walls of blood vessels. With hydroxyproline deficiency the collagen is defective and localized rupture of blood vessels occurs.

2 There are no cross links, which makes the fiber much weaker. The lack of cross-linking is a result of a nitrile in the sweet pea seed that inhibits the enzyme that forms these bonds.

3 The amino acid sequence of histones is nearly the same in all organisms along the evolutionary scale, suggesting that functional histone is extremely intolerant of amino acid changes and that all changes are probably lethal.

4 (a) None. Results (1) and (2) test only one type of linkage. Result (3) is independent of the nature of the linkage. (b) Complete digestion with a protease and a nuclease and identification of a component which is neither a free amino acid or a free nucleotide but consists of one amino acid and one nucleotide. One could show that the amino acid and nucleotide are joined by observing co-migration during electrophoresis or chromatography; since amino acids and nucleotides do not bind noncovalently, one can assume that the amino acid and the nucleotide are covalently joined in this complex.

5 Lys-Trp-Lys and Lys-Tyr-Lys are very effective. The symmetric lysines bind to the phosphates and the tryptophan and tyrosine stabilize binding by stacking to adjacent pyrimidines. Lys-Gly-Lys works, but weakly, because there is no stacking stabilization. Lys-Trp-Gly will bind to the DNA only very weakly, if at all, because of the lack of symmetry. Lys-Gly-Gly will have no affinity for DNA.

CHAPTER 8

THE GENETIC MATERIAL

CHAPTER SUMMARY

As is common in most branches of science, the development of molecular genetics has two features: (1) a collection of significant observations, some made by chance and others resulting from further exploration of these chance observations, and (2) theories, some highly intuitive and others based on detailed examination of experimental data. Griffith's early study of bacterial pneumonia led to the discovery that a bacterial species with a stable genotype could acquire the genotype of a heat-killed variant of the same species simply by putting the living and dead bacteria in contact with one another, for example, by injecting both into a mouse. Avery and co-workers tested various components isolated from the bacterium donating its genotype and discovered that the substance responsible for the genotypic change (the transforming principle) was DNA. In so doing, they provided the experimental foundation for what we now know to be true, namely, that DNA is the genetic material of bacteria. Hershey and Chase showed, several years later, that when T2 phage adsorbed to *E. coli*, only DNA was injected into the bacterium and that this DNA provided all of the genetic information necessary for development of progeny phage. This experiment proved that DNA is the genetic material of phage T2 and suggested that DNA is the genetic material of all phages, all bacteria, and all living cells. Over the next few decades the only exceptions to this rule that have been observed are certain plant and animal viruses and a small number of phages, which utilize RNA instead of DNA as the genetic material.

The discovery that different species contain DNA of widely different base composition and the recognition that an enormous number of base sequences are possible in DNA made it possible to see how DNA could contain genetic information for the synthesis of the large number of proteins synthesized by different cells. Crick and co-workers suggested that it was unlikely that a molecule as precious as DNA would be used repeatedly for the synthesis of every polypeptide chain; thus, they proposed that an intermediate molecule would be synthesized once or a small number of times and that this intermediate would be used in protein synthesis. Realizing that RNA could be copied from DNA, preserving the DNA base sequence, they argued for the existence of what they called messenger RNA and formulated what came to be known as the Central Dogma: *DNA makes RNA makes protein.* Several years after this hypothesis was formulated, messenger RNA was isolated from *E. coli.* The question of the mechanisms by which a base sequence of DNA could be translated into an amino acid of a protein remained. It seemed reasonable that the base sequence should be read linearly and that a particular sequence of bases would carry the information for placement of a single amino acid in a polypeptide chain. Simply by counting the number of possible arrangements of four nucleotides, it was noted that bases had to be read in groups of at least three in order to be matched with 20 amino acids. The set of triplets by which DNA bases and amino acids are related (codons) was called the genetic code. Since amino acids do not bind to RNA, Crick further suggested that a set of adaptor molecules should exist; these molecules would contain a binding site for a particular amino acid, probably via covalent binding, and a sequence of three bases able to base-pair with the three coding bases in the messenger RNA (an anticodon). These adaptors were found and termed transfer RNA. Each active transfer RNA molecule contains a covalently bound amino acid. By a specific codon-anticodon interaction the transfer RNA molecules are aligned along a messenger RNA molecule (two transfer RNA molecules at a time) and the amino acids they carry become covalently linked and ultimately separated from the transfer RNA.

DNA has many features that make it useful as a genetic material: (1) It can store genetic information in the form of its base sequence. (2) It can transfer its information to the protein-synthesizing apparatus of a cell, as needed, by synthesis of an RNA molecule whose sequence is complementary to that of the DNA. (3) It has great physical and chemical stability. The presence of 2'-deoxyribose makes the phosphodiester bonds quite resistant to hydrolysis. The hydrophobic nature of the bases causes them to be in the central portion of the DNA molecule and to have limited contact with water, which reduces attack by reactive molecules. (4) DNA can transfer genetic information to progeny cells by a replicative mechanism in which each strand serves as a template for synthesis of a complementary strand. (5) Genetic change can occur without loss of parental information; this can occur via mutational alterations that result in the production of one daughter molecule identical in base sequence to the parent molecule and a second daughter molecule with a slightly different base sequence.

One difference between DNA and RNA is the use, in DNA, of the pyrimidine thymine (T) to pair with adenine (A), and, in RNA, of uracil (U). This difference probably evolved as a means of maintaining genetic information, because the base cytosine (C) is frequently deaminated to form U. If a DNA molecule in which a C-to-U change had occurred were to replicate, a parental GC pair would be replaced by a mutant AT pair in one of the daughter molecules and information would be lost. However, by having thymine in DNA a successful repair mechanism has been able to evolve that removes all uracil from DNA and replaces it with thymine.

Evolution demands that mutations occur, but at a low frequency. A feature of DNA that allows mutations to occur at low frequency is the existence of tautomeric forms of adenine and thymine. These forms cause infrequent errors in replication.

BOLD TERMS

anticodon, blendor experiment, chemical alteration, codon, genetic code, messenger RNA, mRNA, mutation, replication, replication error, transfer experiments, transfer RNA, transformed, transforming principle, triplets, tRNA

DRILL QUESTIONS

1 What term describes the base sequence that corresponds to a single amino acid?

2 Why is an adaptor molecule needed to read the genetic code?

3 Why is a messenger RNA a valuable intermediate in protein synthesis?

4 What is meant by bacterial transformation?

5 What experimental result showed that transformation consists of a permanent genetic change?

6 What argument was used by Avery and co-workers to rule out protein or RNA as the genetic material?

7 What experiment showed that the role of transforming principle was not simply to cause synthesis of a new capsular polysaccharide?

8 What data eliminated the idea that DNA was a repeating tetranucleotide?

9 In the blendor experiment what radioactive atoms were used to label DNA and protein and why were they chosen?

10 In what organisms is RNA the sole genetic material?

11 Why is RNA but not DNA hydrolyzed at alkaline pH?

12 What is the product of deamination of cytosine?

13 How are uracil and thymine chemically related?

14 What base pairs can be formed by tautomeric forms of adenine and thymine?

ANSWERS TO DRILL QUESTIONS

1 A codon.

2 Amino acids do not bind directly to RNA; if they did, their van der Waals surfaces would not be sufficiently different to be bound to a specific triplet of bases.

3 DNA does not have to be used every time a polypeptide chain is synthesized.

4 The conversion of the genotype of a bacterium to the genotype of a second bacterium by exposure of the first bacterium to DNA from the second.

5 Progeny of a transformed cell all had the new character; in genetic terms, the cells bred true.

6 The ability of a cell extract to transform was not reduced by treatment of the extract with enzymes that depolymerize proteins or RNA.

7 A penicillin-sensitive bacterium could acquire resistance to penicillin by exposure of the cell to DNA from a bacterium resistant to the drug.

8 Chargaff's discovery that many DNA molecules had a base composition in which all bases were not equimolar.

9 ^{32}P for DNA, because there is no phosphorus in phage protein; ^{35}S for proteins, because there is no sulfur in DNA.

10 Many plant and animal viruses and some bacteriophages.

11 The chemical mechanism for hydrolysis of a phosphodiester bond requires participation of a free OH group, which is present on the 2' carbon in RNA but not in DNA.

12 Uracil.

13 Thymine is 5-methyluracil.

14 The adenine tautomer can pair with cytosine; the thymine tautomer can pair with guanine.

PROBLEMS

1 An early structural model for DNA was the so-called tetranucleotide hypothesis: four nucleotides (one each of adenine, cytosine, guanine, and thymine) were thought to be covalently linked to form a planar unit. It was suggested that units are linked together to yield a repeating polymer of the tetranucleotide. Explain why, if DNA had such a structure, it would probably not have become the genetic material of cells?

2 Which of the following are true? The original experiments on transformation in *Pneumococcus* showed that (1) a hereditable character of a bacterium could be permanently altered by exposure to DNA from a bacterium having a different character; (2) a nonheritable character of a bacterium could be permanently altered by exposure to DNA from a bacterium having a different character; (3) all genetic material must contain DNA; (4) all of the genes in a bacterium must be carried on DNA molecules.

3 Following publication of the transformation experiments of Avery, MacLeod, and McCarty, opponents of the DNA=gene theory, who believed that genes were made of proteins, argued that the transformation was caused by proteins that were contaminating the DNA sample. (a) If transformation was indeed carried out by protein rather than DNA molecules and if the DNA preparation used contained at most 0.02% protein, how many protein molecules (each consisting of about 300 amino acids) would have been present in one milliliter of a DNA solution at a concentration of 10^{-7} mg/ml? (b) If protein was the active agent in transformation, would the number calculated in part (a) account for the fact that in a typical transformation experiment 1000 transformants result from 0.0001 g of *Pneumococcus* DNA?

4 In the Hershey-Chase experiment what is the significance of the fact that phage are produced by the infected bacteria after blending?

5 Is it possible in a Hershey-Chase experiment with a newly discovered phage that no ^{32}P would appear in any progeny phage? Explain your answer.

6 You wish to prove that DNA but not protein is the genetic material for a particular phage. You conceive of an experiment like the blendor experiment but unfortunately the phage has a short tail and cannot be removed from the cell wall with a blendor. You know how to isolate and purify progeny phage. You prepare phage labeled with both ^{32}P and ^{35}S, infect bacteria, and collect the progeny from the infection. Which of the following results could be taken as evidence to support your belief: (1) ^{35}S never appears in progeny phage; (2) ^{32}P always appears in several progeny phage but the amount per phage is much less than the amount in the parental phage particles, and ^{35}S appears in no phage particles; (3) ^{32}P appears in only one or two progeny phage particles and in only one strand of the DNA, and ^{35}S appears in many progeny phage but the amount per phage is much less than the amount in the parental phage particle.

7 Suppose you were given two tubes, one with ^{32}P-labeled DNA and the other with ^{32}P-labeled DNA. You mix them up but need them very soon in an experiment.

You know that (1) both can be precipitated with acid and the precipitate can be collected on a filter and detected by their radioactivity, and (2) that bases, nucleosides, and nucleotides are soluble in acid. What simple test might you carry out with a portion of each to distinguish the DNA from the RNA? You do not have any enzymes available, but you have a set of common chemical reagents.

ANSWERS TO PROBLEMS

1 As a simple repeating polymer consisting of a single unit, it cannot carry information for an amino acid sequence.

2 Only statement 1.

3 (a) The solution containing 10^{-7} mg/ml DNA would contain 0.02 percent of 10^{-7} mg = 2×10^{-14} g protein. The molecular weight of a protein containing 300 amino acids is 300 times the average molecular weight of an amino acid, or approximately (300) (100) = 3×10^4. Thus, the number of protein molecules is $(6 \times 10^{23})(2 \times 10^{-14})/(3 \times 10^4) = 4 \times 10^5$. (b) The numbers do not exclude the possibility that the transformation is protein-mediated, since the maximum estimate of the number of protein molecules (4×10^5) exceeds the number of transformants 400-fold.

4 The ^{35}S-containing parts of the phage are removed from the bacterium, and phage production is still possible. Thus they are no longer needed for phage production.

5 Yes, if only progeny, and never parental, DNA is packaged.

6 (1) would not, because with this phage parental genetic material might rarely be packaged. (2) is not very convincing, because the amount of ^{32}P is so small. (3) is relatively convincing.

7 Treat each with hot alkali and then neutralize and precipitate with acid. The RNA will be totally hydrolyzed to nucleotides and will not precipitate; the DNA is resistant to the alkali and remains precipitable. Thus, if the acid-treated samples are filtered, radioactivity will be found with the DNA, but not the RNA, sample.

MORE CHALLENGING PROBLEMS

1 Suppose you wish to prove to a skeptic that transformation is mediated by DNA and not by protein. You do not have available pure enzymes that degrade DNA or protein. However, you have worked hard and have shown (i) that 50 different genetic traits can be transformed, and (ii) that transformation is very inefficient and highly dependent on DNA concentration. The only piece of nonmicrobiological laboratory equipment you possess is an ultracentrifuge and with it you are able both to measure the molecular weight of DNA and fractionate DNA according to molecular weight. Design a simple experiment to prove that the transforming principle is DNA. Hint: Think about linkage.

2 A critic of the interpretation of the transformation experiment might say that protein is the genetic substance and the protein can penetrate the cell only when the protein is bound to DNA. Thus, the loss of transforming activity following boiling of the DNA might be a result of dissociation of protein from the DNA. Assume that you have current knowledge about the denaturation of proteins and of DNA, about the effect of low and high pH on the chemical and physical properties of DNA and protein, and about the ionic strength dependence of the binding of protein to DNA. Design an experiment to prove that the critic is incorrect.

3 Molecular biology was in a formative stage when Avery and his colleagues performed the transformation experiment. Many kinds of instruments and techniques are now available that would have simplified their investigation. Suggest a way that CsCl density gradient centrifugation might have been used to determine whether DNA or protein is the genetic material.

ANSWERS TO MORE CHALLENGING PROBLEMS

1 Isolate DNA and centrifuge it in a sucrose gradient. Fractionate the centrifuge tube and test each sample for transformability using a very low DNA concentration. Try to find two genetic markers that co-transform, that is, two for which the recipient bacterium can acquire both markers even when the DNA is very dilute. You will observe that the genetic element that carries linked markers always sediments more rapidly than the bulk of the elements that carry only one marker. If this can be shown for several pairs of markers, you will have shown that transformability sediments with DNA and that co-transformability requires DNA of higher molecular weight. This should be pretty convincing. You could then also break the DNA molecules (perhaps by hydrodynamic shear) and show that as the average molecular weight of the DNA decreases, there is no loss of transformability of individual markers but that linkage is lost.

2 Show that transformability as a function of pH follows curves for DNA denaturation rather than protein denaturation. Show that transformation occurs at an ionic strength (e.g., 0.3 M) at which DNA-protein interactions are very weak. Separate protein and DNA by gel electrophoresis or gel chromatography, and test purified fractions for transformability.

3 Isolate the DNA and centrifuge to equilibrium in buoyant CsCl. Since DNA and

protein have different densities, they will come to equilibrium at different positions in the gradient. Fractionate the centrifuge tube. The fractions containing DNA only or protein only can be tested separately for transforming activity.

CHAPTER 9

DNA REPLICATION

CHAPTER SUMMARY

DNA replication is accomplished by an enzymatic polymerization reaction in which a DNA strand is used as a template to synthesize a copy whose base sequence is complementary to the template. In the most common form of replication of a double helix each strand serves as a template. Replication is semiconservative in that each parental single strand is present in one of the double-stranded progeny molecules. DNA replicates using enzymes called DNA polymerases to form the phosphodiester bonds.

Other proteins are also used in particular stages of replication. The growth of a DNA strand proceeds by a DNA polymerase (1) bringing in a nucleoside triphosphate capable of hydrogen-bonding with a base in a template strand, (2) removing the terminal diphosphate from the nucleoside triphosphate, and (3) joining the 5'-P group of the nucleoside monophosphate to a free 3'-OH group on the terminus of the growing strand. The polymerases can only join a 3'-OH group and a 5'-P group. The growing daughter strand always carries a 3'-OH group and the source of the 5'-P is always the incoming nucleotide. Since double-stranded DNA is antiparallel, only one strand, the leading strand, grows in the direction of movement of the replication fork. Since the leading strand forms by using one of the parental strands as a template, the other parental strand becomes single-stranded. Periodically, initiation occurs on this strand, but always in the appropriate direction. Thus, the second daughter strand, the lagging strand, is synthesized in the opposite direction as short fragments (precursor fragments) that are subsequently joined together; the overall direction of joining of the fragments is in the direction of movement of the replication fork.

DNA polymerases cannot initiate synthesis, so a primer is always needed. A primer may be any short oligonucleotide that can base-pair with the template strand. However, in most replication modes the primer is a short RNA oligonucleotide synthesized either by the usual RNA polymerase, which is used for the synthesis of most RNA, or by a special RNA polymerase called DNA primase, which only makes RNA primers. The RNA primer remains hydrogen-bonded to the template strand, and the DNA chain is formed by joining deoxynucleotides to the 3' end of the primer. The primer is removed at later stages of replication by the 5'-3' exonuclease activity of DNA polymerase I (this enzyme always acts immediately ahead of the growing strand, if bound nucleotides are present). The leading strand and each fragment of the lagging strand are initiated by RNA primers, but details of the initiation events differ for the two strands.

In *E. coli* (and probably in most bacteria) two polymerizing enzymes, polymerase I and polymerase III, carry out nucleotide addition. Pol III is responsible for overall chain growth; pol I is used in the completion of precursor fragments and also in repair processes. The two enzymes differ significantly in their ability to use various types of molecules as templates. For example, *in vitro* a DNA molecule having one single-strand break with a 3'-OH group can serve as a site of nucleotide addition by pol I but not by pol III. When pol I acts at such a site, two modes of synthesis are possible: (1) It can remove the deoxynucleotides in its path by means of its 5'-3' exonuclease activity (the one needed to remove RNA from precursor fragments prior to joining); this reaction is called nick translation. (2) It displaces the DNA ahead of it, producing a single-stranded tail; this reaction is called strand displacement and is indicative of the ability of pol I to unwind the helix. Pol III fails to carry out both reactions, because it lacks significant 5'-3' exonuclease activity and it is unable to unwind DNA. *In vivo*, pol III always follows behind a helicase, an enzyme that unwinds double-stranded DNA. Because of the precursor-fragment mechanism, synthesis of the lagging strand is always delayed with respect to the leading strand, and hence single-stranded DNA is always present in the replication fork. The single-stranded segment is the unreplicated portion of the template for the lagging strand. Intramolecular hydrogen bonds can form between bases within a single strand, so this single-stranded DNA can collapse on itself as a tangle of partially double-stranded segments. Such an event is prevented by the SSB protein, which binds to single-stranded DNA. In fact, as the helicase unwinds the DNA molecule ahead of the replication fork, SSB protein immediately binds to the DNA to prevent formation of double-stranded DNA. Pol III is able to displace the SSB protein, and replication proceeds.

Pol I and pol III both possess a 3'-5' exonuclease activity that is used for correcting incorporation errors. If a nucleotide that cannot form correct hydrogen bonds with the template is added to the end of the growing strand, the polymerase undergoes a conformational change that activates the 3'-5' exonuclease. The polymerase then moves backwards and removes the unwanted nucleotide. Since the new terminal nucleotide is correctly base-paired, the polymerase returns to its normal form and polymerization ensues. As mentioned above, pol I also possess a 5'-3' exonuclease activity that can remove one or more nucleotides sequentially ahead of the enzyme. This activity is directed against both ribosides and deoxyribosides and is used *in vivo* to remove the RNA primer from precursor fragments prior to joining these fragments.

Replication of both linear and circular DNA molecules is usually bidirectional in that the replication loops enlarges by movement of replication forks at both ends of the loop. A DNA molecule of a prokaryote is usually initiated at a single site and hence has a single replication loop; in contrast, eukaryotic DNA molecules are multiply initiated and have many loops. Circular molecules replicate in one of two ways: in the θ form or by rolling circle (σ) replication. The latter generates a linear branch, which may be either single- or double-stranded. In the θ mode DNA gyrases are needed to prevent tangling of the unreplicated portion.

Termination of a round of DNA synthesis occurs in several ways. In circular bidirectionally replicating molecules termination usually occurs by collision of the two growing forks at whatever point in the DNA molecule at which the forks happen to collide; with only few exceptions the termination site is not determined by any particular base sequence. The result of termination by collision is a pair of interlocked circles, called a catenane. In *E. coli* the circles are separated by DNA gyrase and, in other systems, by similarly acting topoisomerases. With a few plasmids, termination has been observed to occur at defined nucleotide sequences. Termination of linear molecules is poorly understood. With *E. coli* phage T7, DNA molecules that have replicated except for their termini are joined via their complementary single-stranded termini to form a linear dimer, which then initiates a second round of replication. The dimer also replicates without duplicating the termini and forms a tetramer, and so forth. Ultimately, these concatemeric units are cut into monomers. For completion of the termini of the linear DNA molecule of a eukaryotic chromosome two terminal regions called telomeres are required. Telomeres contain repeated nucleotide sequences, single-strand breaks, short gaps, and often terminal hairpins. The mechanism by which the telomere enables termination to occur is unknown.

BOLD TERMS

AP endonuclease, bidirectional replication, catenation, catenane, concatemer, conservative replication, core enzyme, covalent extension, D loop, daughter strand, *de novo* pathway, denaturation mapping, displacement loop, DNA gyrase, DNA ligase, DNA polymerase, dUTPase, editing function, helicase, hybrid DNA, initiation, kinases, lagging strand, leading strand, looped rolling-circle replication, mismatch repair, nick translation, node, nucleoplasmin, nucleoside diphosphate kinase, Okazaki fragments, parent strand, pol I, pol III holoenzyme, polymerase I, preprimosome, primase, primer, primosome, proofreading, pulse-chase experiment, pulse-labeling experiment, pyrophosphatase, Rep protein, replication origin, rifampicin, RNA polymerase, ribonucleotide reductase, rolling-circle replication, salvage pathway, semiconservative replication, σ replication, sign inversion mechanism, SSB protein, template, terminal redundancy, thymidylate synthetase, topoisomerase, uracil *N*-glycosylase

ADDITIONAL INFORMATION

In the text it was pointed out that in *E. coli* both a helicase and SSB protein are needed for advance of the growing. A requirement for two proteins is not universal in that in some systems a single protein possesses both activities. For example, in the replication of the DNA of phage T4 a phage-encoded protein, the 32-protein (product of gene *32*), serves both functions. The 32-protein is a DNA-binding protein that binds to single-stranded DNA and to itself so tightly that it can unwind DNA (in fact, *in vitro* it can cause complete denaturation of a linear DNA molecule). Thus, instead of fueling the unwinding by ATP hydrolysis, as is the case with SSB protein, the binding energy between the individual protein molecules provides the energy for unwinding. T4 also makes its own DNA polymerase, which displaces the 32-protein during polymerization.

One of the major unsolved problems of *E. coli* DNA replication is the mechanism of initiation of a round of replication. In 1961 it was shown that inhibition of protein or RNA synthesis prevents initiation, which suggested that synthesis of a protein is required. We know now that transcription is also necessary in order to provide a primer. A model that was proposed some years ago hypothesized that initiation might be determined by the concentration of a key protein: as the cell enlarges, the concentration of the protein decreases until at a critical low concentration initiation would occur. This idea was in part a result of the observation that, in *E. coli,* initiation of replication occurs near the middle of the life cycle of each cell and not immediately following cell division. It has been known for some time that a critical element is the DnaA protein, which acts in an unknown way on *oriC,* the replication origin. Recent experiments have shown that DnaA regulates its own synthesis in that when its concentration drops below a critical value, there is a burst of transcription of the *dnaA* gene and hence of synthesis of the DnaA protein. Other experiments show that if the amount of DnaA protein is artificially increased in the cell, well above a certain value, multiple initiation events can occur. These observations suggest that as the cell enlarges, the concentration of DnaA protein decreases until synthesis of DnaA protein is stimulated. However, careful examination of these and other experiments indicate that an increase in the concentration of DnaA protein cannot be the sole determinant of initiation, and that several factors, as yet unknown, are involved.

The mental exercise of speculating about why a system evolved as it has is often useful. Let us consider why an RNA primer rather than a DNA primer might have evolved. When raising a "why" question, it is important to think about the problems faced by a system and the consequences of an alternative mechanism. We have seen that great effort goes into maintaining high fidelity of replication. It seems likely that if an error is made in the first few bases of a primer, an editing function that detects unpaired bases might not be able to correct it; this is because stacking provides little stabilization of the double-stranded structure formed by a very short oligonucleotide, so hydrogen-bonding is very weak. Thus, the error frequency probably cannot be reduced to an acceptable level in the initial portion of the primer by editing. It may also be the case that, for a variety of reasons, priming is an error-prone process. If there is no good way to correct these errors, an effective solution is to discard the primer and recopy the DNA segment by a minimal-error process. This is, of course, exactly what is done by the 5-3' exonuclease of pol I. Thus, the advantage of using RNA as a primer might be that it is easily recognized by pol III (it is chemically different from DNA) as a signal to stop synthesis so that pol I can nick-translate; nick-translation is, of course, a minimal-error process because the 3'-5' editing function remains active.

DRILL QUESTIONS

1 Is DNA replication conservative or semiconservative?

2 Fill in the blank word. A parent strand serves as a _____ for synthesis of a daughter strand.

3 In semiconservative replication what fraction of the DNA consists of one parental strand and one daughter strand after 1, 2, and 3 rounds of replication?

4 In which mode of replication does a parental circular DNA molecule yield two daughter circles?

5 In which mode of replication does a parental circle generate a circle with a linear branch?

6 Will a ^{15}N-labeled circle replicating in ^{14}N medium using the rolling circle mode ever achieve the density of ^{14}N^{14}N DNA?

7 How does rolling circle replication differ from θ replication with respect to the templates that are used?

8 What is the sense of the supercoiling produced by replicative movement of the growing fork in a circle?

9 What is the sense of supercoiling produced by DNA gyrase?

10 Can DNA gyrase convert a positive supercoil to a negative supercoil?

11 Which enzyme is able to separate two catenated circles?

12 Name three enzymatic activities of DNA polymerase I.

13 In what direction does a DNA polymerase move along a template strand?

14 What are the precursors for DNA synthesis?

15 Why is a pyrophosphatase necessary in DNA polymerization reactions?

16 What is the immediate precursor of TMP?

17 DNA polymerization occurs by addition of a deoxynucleotide to what chemical group?

18 What are the roles of the various exonuclease activities of the DNA polymerases in DNA replication?

19 What two reactions are coupled when nick translation occurs?

20 What is the difference in the activity of pol I and pol III at a single-strand break with a free 3'-OH group?

21 If pol I acts at a single-strand break without the 5'-3' exonuclease activity, what happens?

22 How do pol I and pol III differ with respect to their ability to unwind the parental DNA in a replication fork?

23 Distinguish the roles of polymerases I and III in DNA replication.

24 What is the role of a helicase in DNA replication?

25 What is the function of SSB protein in advance of the replication fork?

26 Which enzymatic activity of pol I and of pol III is responsible for proofreading?

27 What chemical groups are joined by DNA ligase?

28 What is the chemical difference between the groups joined by a DNA polymerase and by DNA ligase?

29 How do organisms solve the problem that all DNA polymerases move in the same direction along a template strand, even though double-stranded DNA is antiparallel?

30 Are precursor fragments found in the leading or lagging strand?

31 Which strand, leading or lagging, is usually separated from the replication fork by a gap?

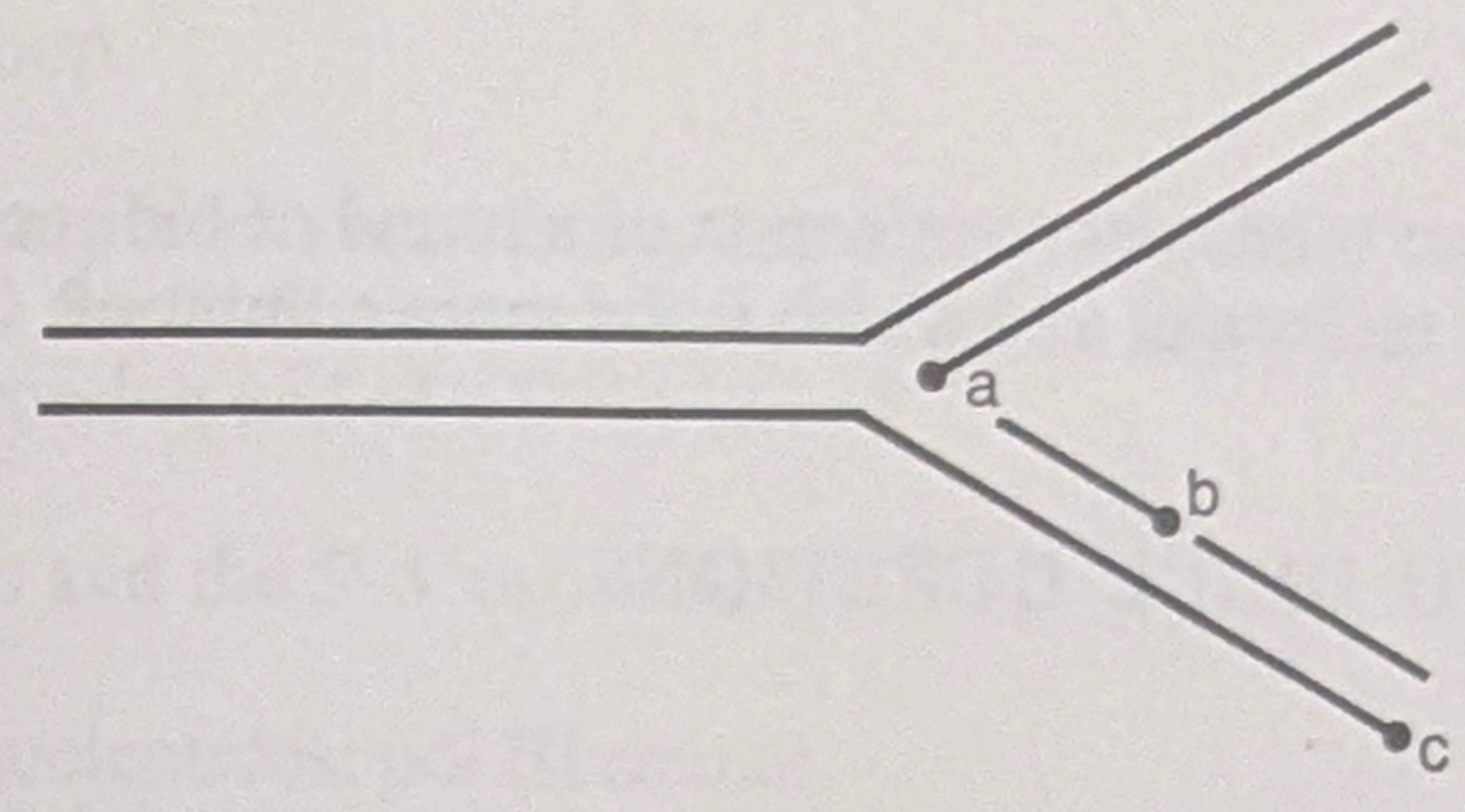
32 Are uracil fragments present in both leading and lagging strands?

33 What is the probable sequence of events used to join two uracil fragments?

34 How is it shown experimentally that small fragments (Okazaki fragments) are present in the region of the replication fork?

35 Which enzymes and in what sequence are used in joining precursor fragments? Assume both fragments are already made.

36 What is the chemical group (3'-P, 5'-P, 3'-OH, 5'-OH) at the sites indicated by the dots labeled a, b, and c in the figure below?



37 An *E. coli* *dut*$^-$ mutant fails to make dUTPase. What effect does this mutation have on the size of Okazaki fragments?

38 Why can RNA polymerases and primases initiate DNA replication whereas DNA polymerases cannot?

39 How does an RNA primer differ from a typical RNA molecule copied from DNA?

40 Distinguish *E. coli* RNA polymerase and primase with respect to sensitivity to rifampicin.

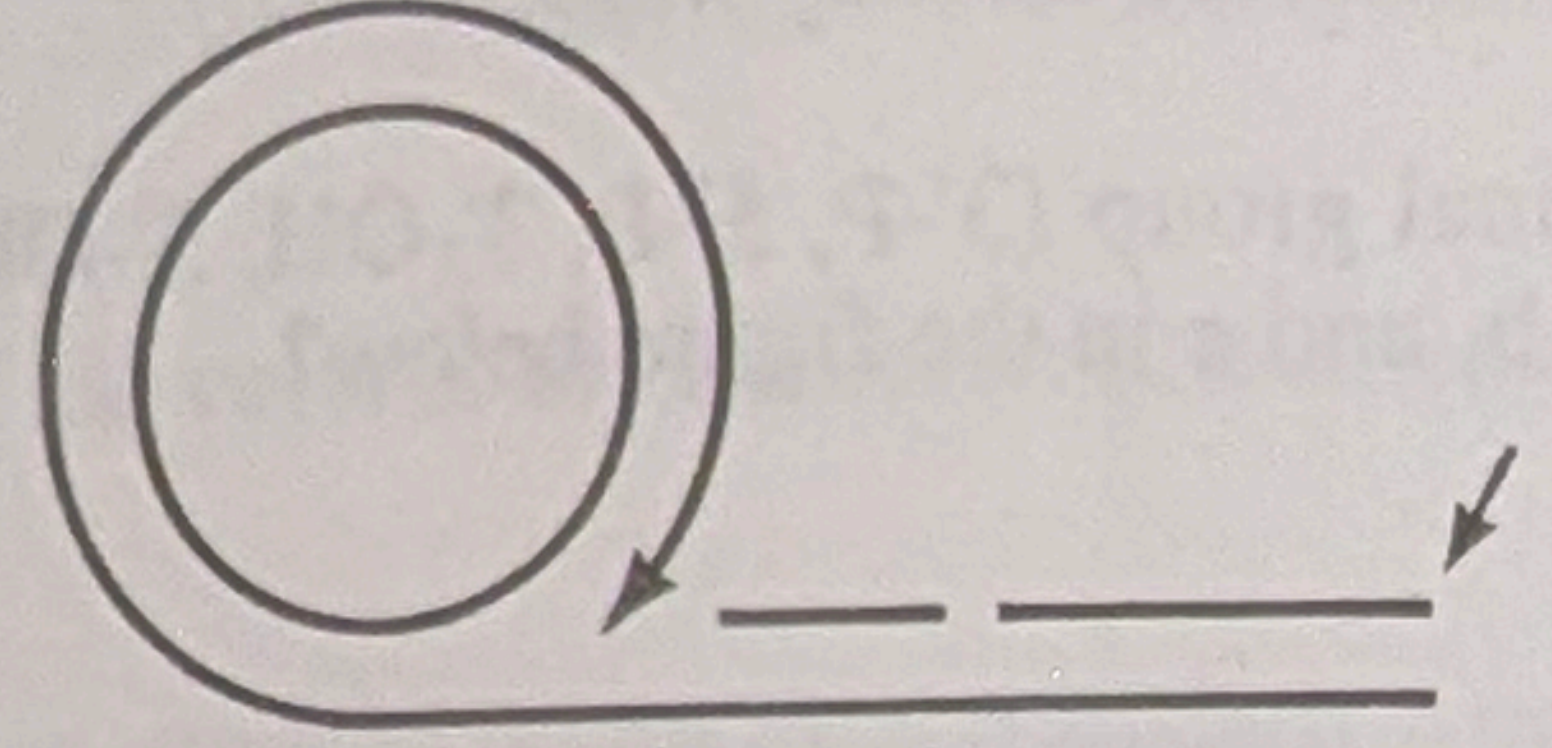
41 What must be done to two precursor fragments before they can be joined together?

42 Which enzymatic activity of pol I is responsible for removal of RNA primers of precursor fragments?

43 What is the molecular difference between the preprimosome and the primosome?

44 What is the fundamental difference between the initiation of θ replication and of rolling circle replication?

45 What is the chemical group that is at the indicated terminus of the daughter strand of the extended branch of the rolling circle shown below?



46 What are the two modes of termination of a round of bidirectional replication of a circular DNA molecule, and which is the more common?

ANSWERS TO DRILL QUESTIONS

1 Conservative.

2 Template.

3 All, 1/2, 1/4, respectively.

4 θ replication.

5 Rolling circle replication.

6 No. After one round of replication it has hybrid density but an ^{15}N strand always remains in the circle.

7 In rolling circle replication only one strand of the parental circular molecule serves as a template; the second strand of the branch is copied from a progeny strand. In θ replication both parental strands are templates.

8 Positive supercoiling.

9 Negative supercoiling.

10 Yes.

11 DNA gyrase.

12 Polymerizing activity, 5'-3' exonuclease, 3'-5' exonuclease.

13 From the 3' end to the 5' end.

14 Deoxynucleoside triphosphates.

15 When pyrophosphate is present, even at low concentrations, the preferred direction of the reaction catalyzed by DNA polymerases is hydrolysis, not polymerization; removal of pyrophosphates shifts the equilibrium and prevents the hydrolytic reaction from occurring.

16 dUMP.

17 To a 3'-OH group.

18 The 5'-3' activity removes ribonucleotides from the 5' termini of precursor fragments, and the 3'-5' exonuclease removes a base that has been incorrectly added to the growing end of a DNA strand.

19 Polymerization and the 5'-3' exonuclease.

20 Pol I can add nucleotides; pol III cannot.

21 Strand displacement.

22 Pol III requires a helicase to unwind the DNA; pol I can unwind a helix without any accessory protein.

23 Polymerase III is responsible for the addition of all nucleotides in the growing fork; polymerase I removes RNA from the 5' end of precursor fragments and replaces it with deoxynucleotides.

24 A helicase unwinds the parental helix in the replication fork.

25 It prevents reformation of double-stranded DNA after the parental DNA is unwound by a helicase.

26 The 3'-5' exonuclease activity.

27 A 3'-OH group and a 5'-P group.

28 DNA polymerases join a 5'-*tri*phosphate to a 3'-OH group and in so doing remove two phosphates, so that the phosphodiester bond contains one phosphate; a ligase joins a 5'-*mono*phosphate to a 3'-OH group.

29 One strand is copied from the 3' end to the 5' end, and the other strand is copied in the direction opposite to that of the movement of the replication fork by synthesis in short pieces.

30 The lagging strand.

31 The lagging strand.

32 Yes, but they are solely responsible for fragmentation of the leading strand.

33 Removal of uracil by a glycosylase, removal of the deoxyribose, gap-filling by polymerase I, and then joining by ligase.

34 Pulse-labeled DNA appears in small fragments when sedimented in alkali. If pulsed and chased, the radioactive DNA has the same average size as the bulk of the cellular DNA.

35 Pol I (5'-3' exonuclease activity), pol I (polymerizing activity), DNA ligase.

36 Remember that the group at the growing end of the leading strand is always a 3'-OH group because DNA polymerases can only add nucleotides to such a group. Because DNA is antiparallel, the opposite end of the leading strand is a 5'-P group. Hence: a, 3'-OH; b, 3'-OH; c, 5'-P.

37 They are smaller due to increased incorporation of dUMP and subsequent removal of these nucleotides.

38 Neither RNA polymerases nor primases require a primer.

39 It remains completely hydrogen-bonded to the DNA template.

40 RNA polymerase, inhibited; primase, resistant.

41 The RNA primer must be removed from the precursor fragment that was made first, because DNA ligase cannot join DNA to RNA.

42 The 5'-3' exonuclease.

43 The primosome contains primase.

44 Rolling circle replication must be initiated by a single-strand break. θ replication does not need such a break.

45 Rolling circle replication begins with a break that produces a 3'-OH group and a 5'-P group. The polymerase extends the 3'-OH terminus and displaces the 5'-P end. Since the DNA strands are antiparallel, the strand complementary to the displaced strand must be terminated by a 3'-OH group.

46 Collision of the two growing forks (more common) and cessation of replication at a particular base sequence.

PROBLEMS

1 A culture of bacteria grown for a very long time in ^{14}N-containing medium is transferred to ^{15}N-containing medium for one-half generation. What is the density distribution if the DNA is isolated either in the standard way in the form of fragments having a molecular weight of 5-10 million, or as an intact molecule?

2 In experiments that confirmed that DNA replication is semiconservative, *E. coli* DNA was labeled by growing cells for a number of generations in a medium containing the heavy ^{15}N isotope and then observed the densities of the molecules during subsequent generations of growth in a medium containing a great excess of ^{14}N. If DNA replication were conservative (that is, if parental single strands never separated but remained associated), what pattern of molecular densities would have been observed after one and after two generations of growth in the ^{14}N-containing medium?

3 Assume that *E. coli* cells grown for many generations in a medium containing ^{14}N are transferred for exactly one generation to a ^{15}N-containing medium. The cells are then transferred back to a medium with an excess of ^{14}N. What are the expected density classes of the molecules and the relative proportions of these classes, if the DNA is isolated and examined after one final generation of growth in the ^{14}N-containing medium? After two final generations of growth in the ^{14}N-containing medium?

4 Assume that you have a culture of *E. coli* cells grown for many generations in a ^{15}N-containing medium. The cells are washed and transferred to a ^{14}N-containing medium. After exactly two chromosome replications in this second medium, the DNA is extracted without any breakage whatsoever (usually not possible, but it has been done). What density classes would result and in what proportions?

5 Why is the activity of most polymerases inhibited by the addition of a chelating agent such as EDTA?

6 It has been stated that helicases are needed because most DNA polymerases cannot unwind a double helix. However, before helicases were discovered, enzymologists were able to replicate DNA using reaction mixtures that probably contained no helicases. (a) What property of DNA allows advance of the replication fork without a helicase? (b) What effect would you expect a helicase to have on either the rate or fidelity of replication?

7 It has been pointed out that an intracellular pyrophosphatase prevents degradation of DNA by the hydrolytic activity of DNA polymerases. Polymerization does occur *in vitro* even if no pyrophosphatase is present. Why is pyrophosphatase needed *in vivo* but not *in vitro*?

8 The doubling time of a bacterial culture depends on the carbon source in the growth medium and ranges from about 20 minutes to nearly ten hours at 37 C. However, over a wide range of cell doubling times, the *E. coli* chromosome goes through an entire cycle of DNA replication in 40 minutes. At what average rate (expressed in nucleotides per second) do the replication forks progress along the chromosome?

9 If polymerase I is supplied with dATP and dTTP, but no primer and no template, DNA containing dA and dT will actually form. Interestingly, the base sequence is a strict alternation of A and T, that is, ATATATATAT.... The reaction is exceedingly slow, taking many hours before DNA appears. However, once DNA is detected, the amount then begins to increase very rapidly. Explain this rapid increase.

10 Answer the following questions about rolling circle replication. (a) How many *de novo* initiation events are needed in a complete replication cycle? (b) If the circle has unit length, what is the maximum length of the linear branch that can be formed? (c) What is the primer?

11 If a $^{15}N^{15}N$ circle replicates in ^{14}N medium by the rolling circle mode, what will be the density distribution after three rounds of replication?

12 Rolling circle replication initiates at a nick. Assume that polymerase III is the replication enzyme, discuss several events that must occur before polymerase III can initiate polymerization.

13 Consider a hypothetical phage whose DNA replicates exclusively by rolling circle replication. A phage whose DNA is radioactive in both strands infects a bacterium and is allowed to replicate in nonradioactive medium. Assuming that only daughter DNA from the branch ever gets packaged into progeny phage particles, (a) what fraction of the parental radioactivity will appear in progeny phage, and (b) how many progeny phage will contain radioactive DNA? Will the occurrence of genetic recombination affect this value?

14 Is there any maximum to the length of the branch of a rolling circle?

15 In bidirectional replication does movement of the two replication forks initiate simultaneously? Explain.

16 How long will it take to replicate a medium-sized phage DNA molecule whose molecular weight is 25 million (a typical size) and which replicates bidirectionally? You may use the fact that *E. coli* DNA replicates in 40 minutes.

17 The origin of the *E. coli* chromosome is on the genetic map shown in Figure 17. What is the sequence (in time) of replication of the genes *A* through *G*? State any assumptions that you have made.

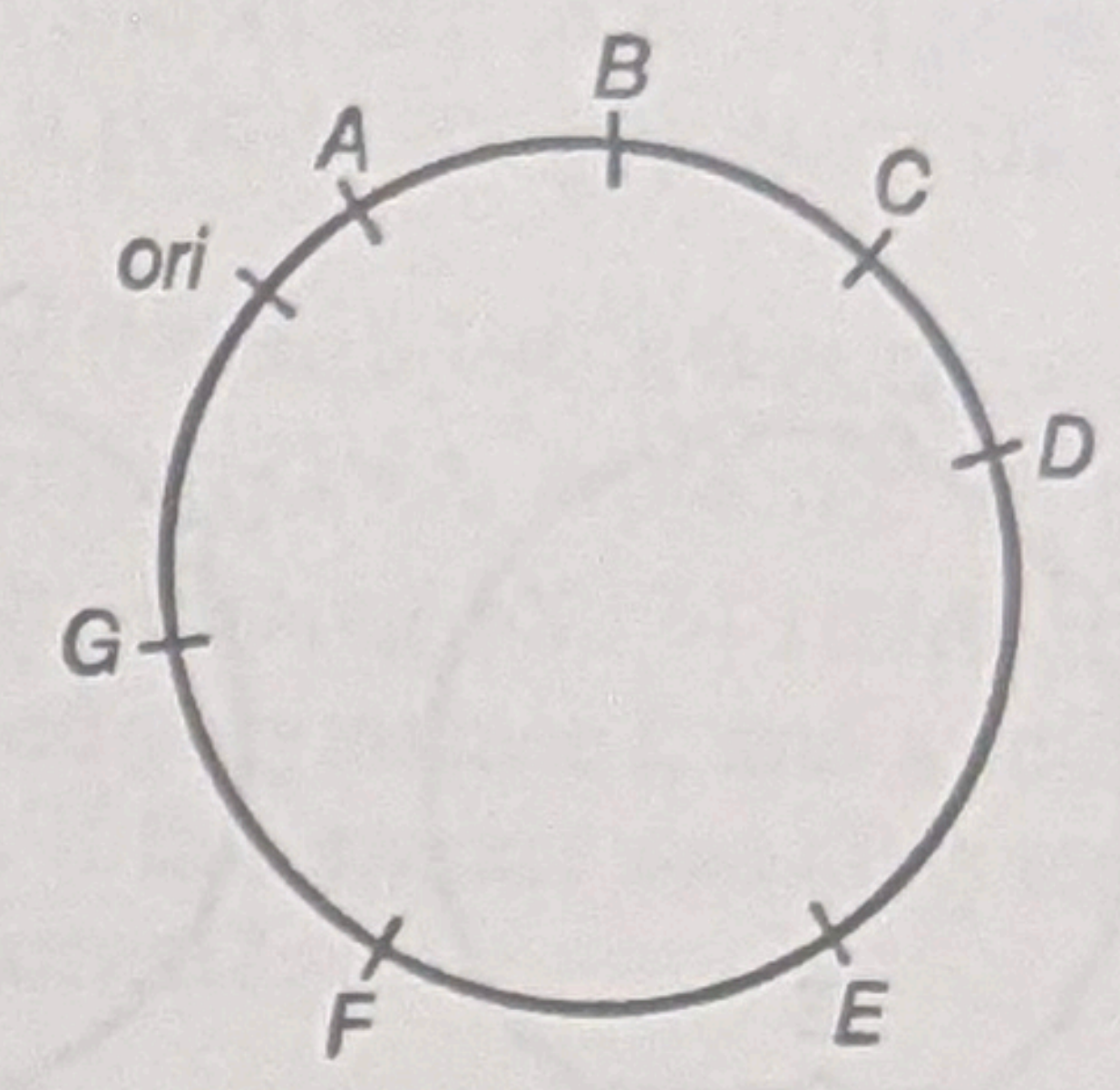
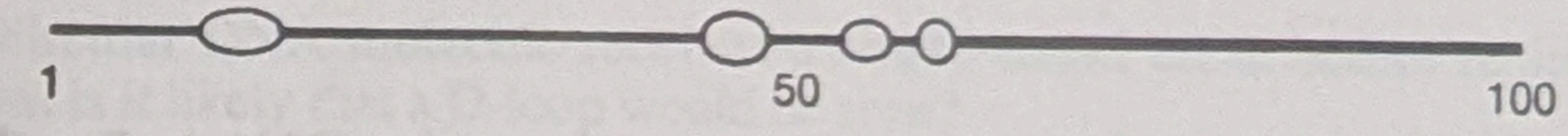
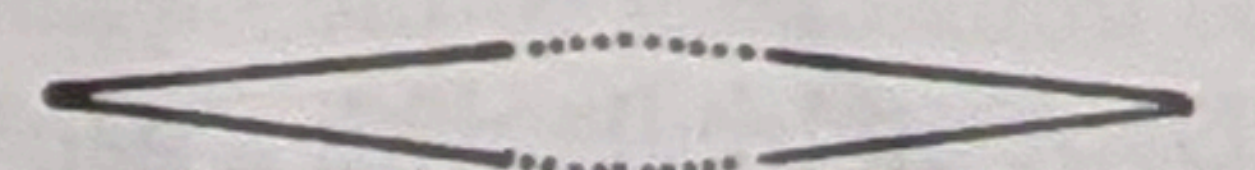


FIGURE 17

18 Bidirectional replication was first detected during vegetative growth of *E. coli* phage λ by electron-microscopic observation of partially denatured DNA. A partially denatured linear DNA isolated from another phage particle has the denaturation map shown in the figure below. Draw a partially denatured, replicating, circular DNA molecule that would be obtained after 50% of the molecule is replicated, if the molecule replicates unidirectionally from a position exactly in the middle of the linear DNA. Repeat for bidirectional replication.



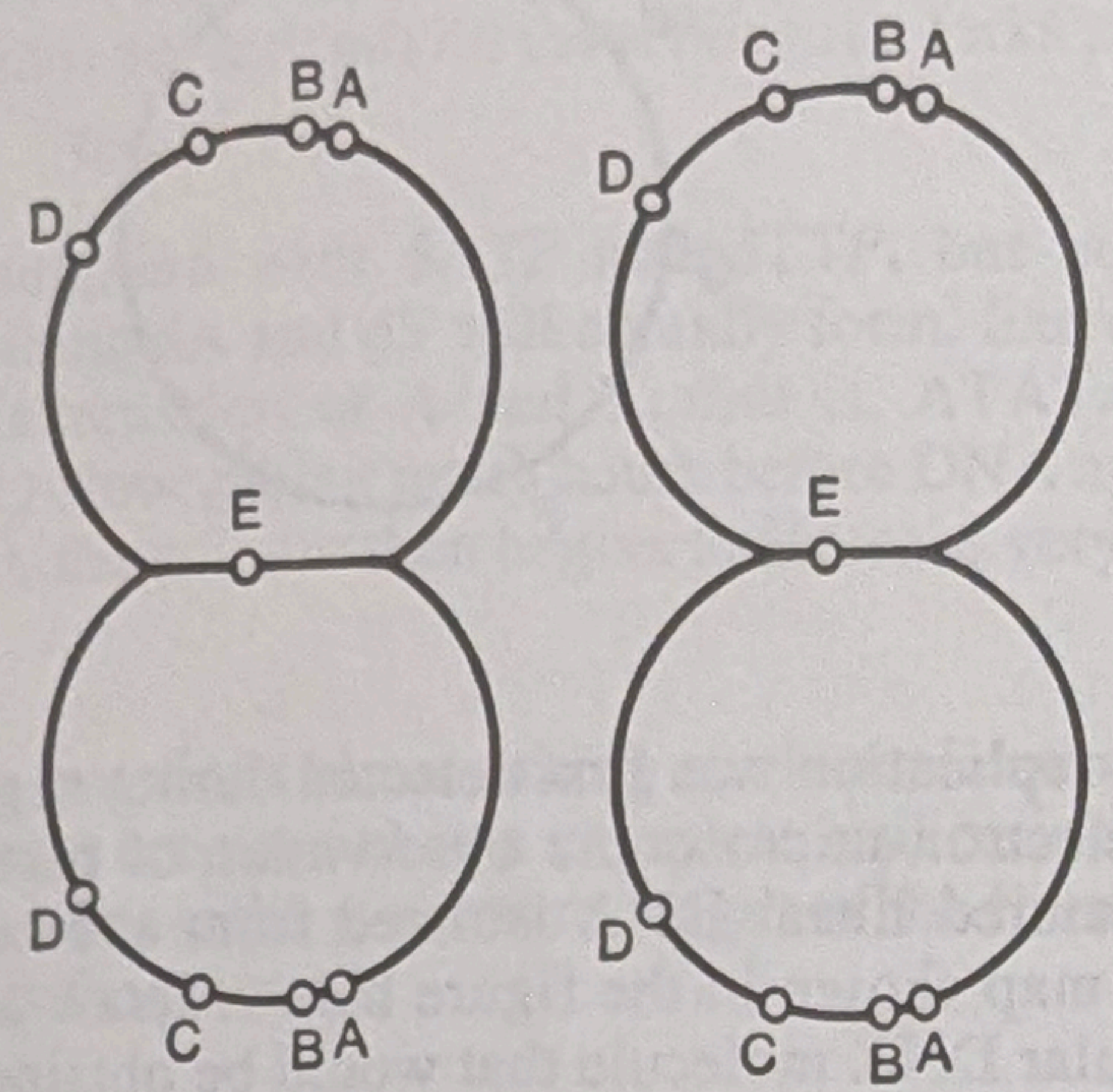
19 Bidirectional replication of eukaryotic DNA has been demonstrated by pulse-labeling with ^{3}H-labeled thymidine and observing the DNA by autoradiography. The result is shown below. (a) How would the autoradiogram appear if replication were unidirectional? (b) What would be the appearance if there were multiple unidirectionally replicating units?



20 What kinds of recognition sites on DNA are required for unidirectional replication, bidirectional replication, and replication of a single-stranded circular DNA?

21 Consider the phage DNA in Problem 18 and assume it replicates bidirectionally. In order to determine whether termination results from simple collision of the replication forks or by stopping of the forks at a genetically defined site, a phage DNA molecule is constructed in which the piece of DNA located 36–47% from

the left end is duplicated in tandem sequence and the piece from 59–68% is deleted. The following molecules were observed. Is there a defined terminus?



80% replicated

90% replicated

22 The gene *dnaE* codes for the polymerase that makes DNA in *E. coli.* The only known *dnaE*⁻ mutants are temperature-sensitive; that is, replication is normal at 30 C, but at 42 C, there is no DNA synthesis. (a) Assume a culture of the cells is pulse-labeled with ^{3}H-thymidine at 30 C. Then, the ^{3}H-thymidine is removed from the growth medium and at the same time the cells are shifted to 42 C. What happens to the precursor fragments during the period at 42 C? (b) Would you expect to find radioactive precursor fragments in a *dnaE*⁻ mutant grown at 42 C in the presence of ^{3}H-thymidine?

23 Suppose you have a DNA molecule with a gap in one strand 5000 nucleotides long and terminated with a 3'-OH group and a 5'-P group. If a DNA polymerase is added to this molecule *in vitro* (with everything else needed to make DNA), will the DNA filling the gap be a single piece or consist of short fragments? Would you expect fragments of any kind if the gap was filled *in vivo*?

24 In early studies of the properties of Okazaki fragments, it was shown that these fragments could hybridize to both parental strands of *E. coli* DNA. This was taken as evidence that they are synthesized on both branches of the replication fork. Even without any rigorous evidence contradicting the idea that precursor fragments are found in both daughter strands, a particular characteristic of the replication of *E. coli* DNA invalidates a conclusion based solely on the hybridization data. What is that characteristic?

25 What experimental result shows that in *E. coli* primase rather than the principal RNA polymerase makes a primer for precursor fragments?

26 An exciting experiment of the mid-1960s was the enzymatic conversion *in vitro* of the single-strand circular DNA molecule of phage φX174 to a double-stranded covalent circle by mixing φX174 DNA, the four deoxynucleoside triphosphates, highly purified polymerase I, and a cell extract that had been boiled and contained no protein. If this extract was dialyzed, the polymerization did not occur. What was in the extract?

27 Naturally occurring supercoiled DNA molecules have about one negative twist per 200 base pairs. When ethidium bromide is added to a sample of supercoiled DNA, binding occurs, but once the number of positive twists of the DNA reaches about one twist per 150 base pairs no more ethidium bromide can bind. (a) What fraction of the molecular weight of a naturally occurring supercoil can be replicated without the help of DNA gyrase? (b) How does this fraction compare to the size of a D-loop in mitochondrial DNA? Can a DNA molecule with positive twists be replicated?

28 Would a D-loop be seen in a linear DNA molecule?

29 If a circular DNA molecule receives a single-strand break before replication begins, is it likely that a D-loop would be seen?

30 DNA polymerases form phosphodiester bonds between 3'-OH and 5'-P groups. Since the joining reaction requires an activation energy,, the polymerases use nucleoside triphosphates containing a high-energy phosphate bond. However, DNA ligases are capable of joining together a 3'-OH and 5'-P group. How do the ligases satisfy the requirement for activation energy?

31 In the interest of economy, cells do not usually make more of any enzyme than is needed. What do you think would be the relation between the number of DNA polymerase III molecules in *E. coli* and the corresponding polymerase in animal cells?

ANSWERS TO PROBLEMS

1 After one-half generation half of the DNA is unreplicated and will still have the original density. The other half has replicated, yielding hybrid DNA, but the mass of the replicated material is twice that of the unreplicated region. Therefore, if fragmented, 2/3 of the DNA will be hybrid and 1/3 will have the parental density. If unfragmented, there will be a single band whose density is

the average of 2/3 hybrid and 1/3 parental.

2 In conservative replication the parental strands remain together and the daughter molecule consists of two newly synthesized strands. Thus, in a density-shift experiment no hybrid DNA will ever form. After one generation there will be two bands—$^{15}N^{15}N$ (parental) and $^{14}N^{14}N$ (daughter molecule)—in equal amounts. After two generations the same two bands will be present, but the ratio of the amounts is 1:3.

3 One generation after the transfer all of the DNA is hybrid ($^{14}N^{15}N$). Continued growth for one generation yields two bands, $^{14}N^{14}N$ and $^{15}N^{14}N$, in equal amounts. After two generations the $^{14}N^{14}N$ DNA yields two identical $^{14}N^{14}N$ molecules and the $^{15}N^{14}N$ molecules produce one $^{14}N^{15}N$ molecule and one $^{14}N^{14}N$ molecule. Thus, there are two bands (light and hybrid) in a ratio of 3:1.

4 The lack of breakage has no effect after a full generation, so there are equal amounts of $^{14}N^{14}N$ and $^{14}N^{15}N$.

5 Almost all polymerases require the Mg^{2+} ion for enzymatic activity. This is also the case with most nucleases.

6 (a) Breathing. (b) Without a helicase replication depends on breathing as the sole cause of unwinding. Thus, replication is much slower when there is no helicase. There is no effect on fidelity.

7 *In vitro* the concentration of pyrophosphate produced by polymerization is too low to cause the equilibrium to shift toward hydrolysis. *In vivo* there are many intracellular reactions in which ATP is cleaved, so that the pyrophosphate concentration is much higher than in an *in vitro* reaction.

8 The molecular weight of *E. coli* DNA is 2.6×10^9, so there are 7.9×10^6 nucleotides. Complete synthesis occurs in (40)(60)= 2400 seconds, or $7.9 \times 10^6/2400$ = 3296 nucleotides/sec. However, replication in *E. coli.* is bidirectional, so two forks are used to replicate the entire chromosome. Hence, each fork moves at half the rate just calculated, or 1648 nucleotides/second/fork.

9 Obviously the enzyme has great difficulty forming any DNA at all, so a long time elapses before any DNA forms. However, once formed, it can fold back on itself and form a double-stranded hairpin, because of the alternating ATAT... sequence. This double-stranded DNA provides a template. Some of the short dinucleotides that form in the early stages of template-free synthesis can serve as a primer. Thus, as soon as some DNA is made, the rate of synthesis begins to approach that for normal template-mediated replication.

10 (a) One. (b) There is no maximum. (c) One of the parental single strands.

11 After one round of replication, all will be hybrid; after the second round, 2/3 will be hybrid; after the third round, 50% $^{15}N^{14}N$, 50% $^{14}N^{14}N$.

12 Polymerase III cannot replicate from a nick so something must occur to form a gap. However, this situation is no different from that at a growing fork in replication and the combined action of helicase and SSB protein should create the necessary conditions. From what we know about replication, polymerase I should be able to carry out rolling circle replication without polymerase III or helicase, but that is probably not the case.

13 (a) In rolling circle replication one parental strand remains in the circular portion and the other is at the terminus of the branch. Therefore, only half of the parental radioactivity can appear in progeny. (b) If there is no recombination between progeny DNA molecules, one progeny phage will be radioactive. If recombination occurs once, two particles will be radioactive. If recombination is frequent and occurs at random, the radioactivity will be distributed among the progeny.

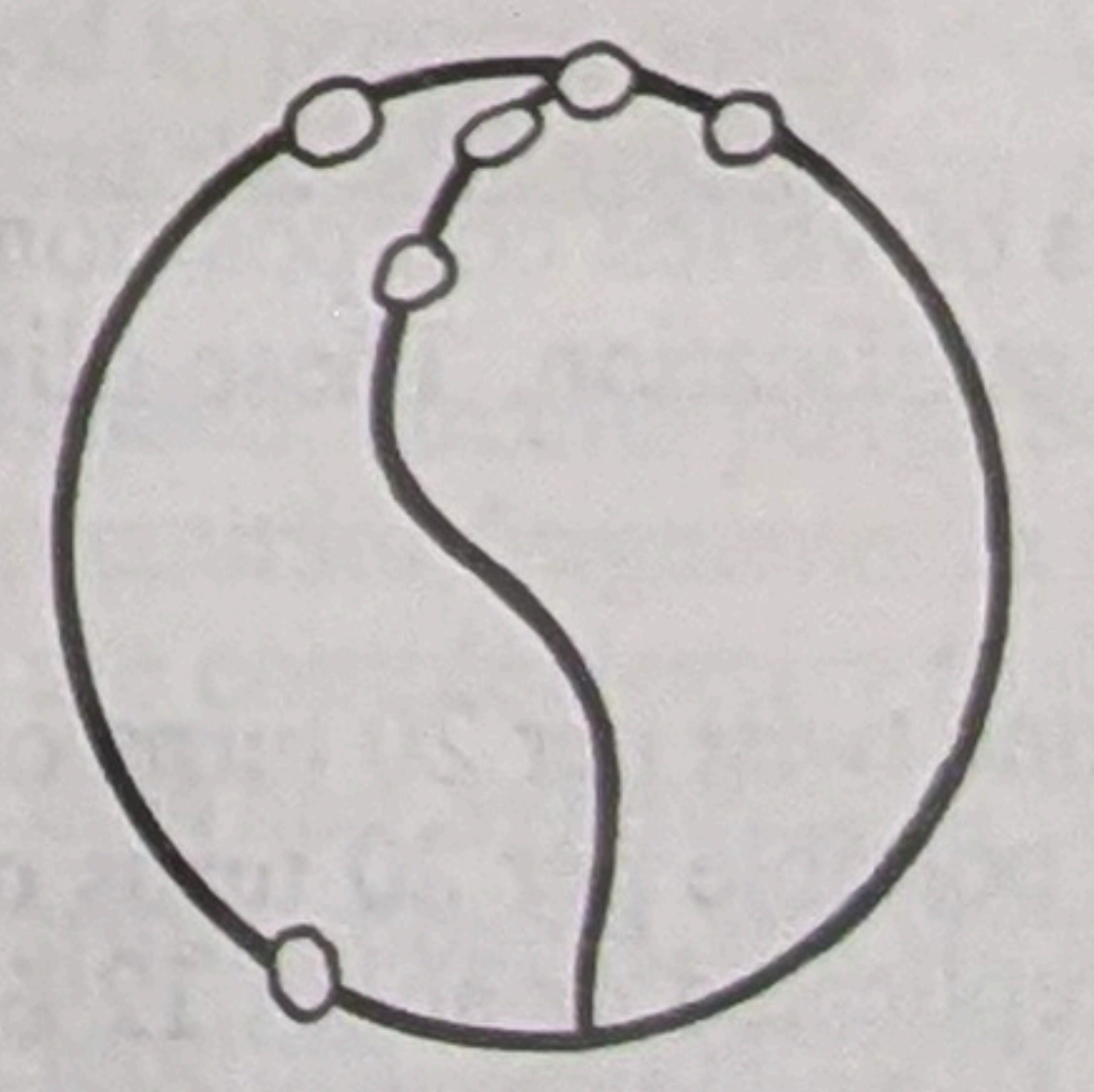
14 There is no maximum, because there is no stop signal for rolling circle replication; the branch grows indefinitely.

15 No. The initiation of the leading strand of one fork is the first event. After a brief time interval synthesis of the lagging strand occurs. The lagging strand is the primer for the second replication fork.

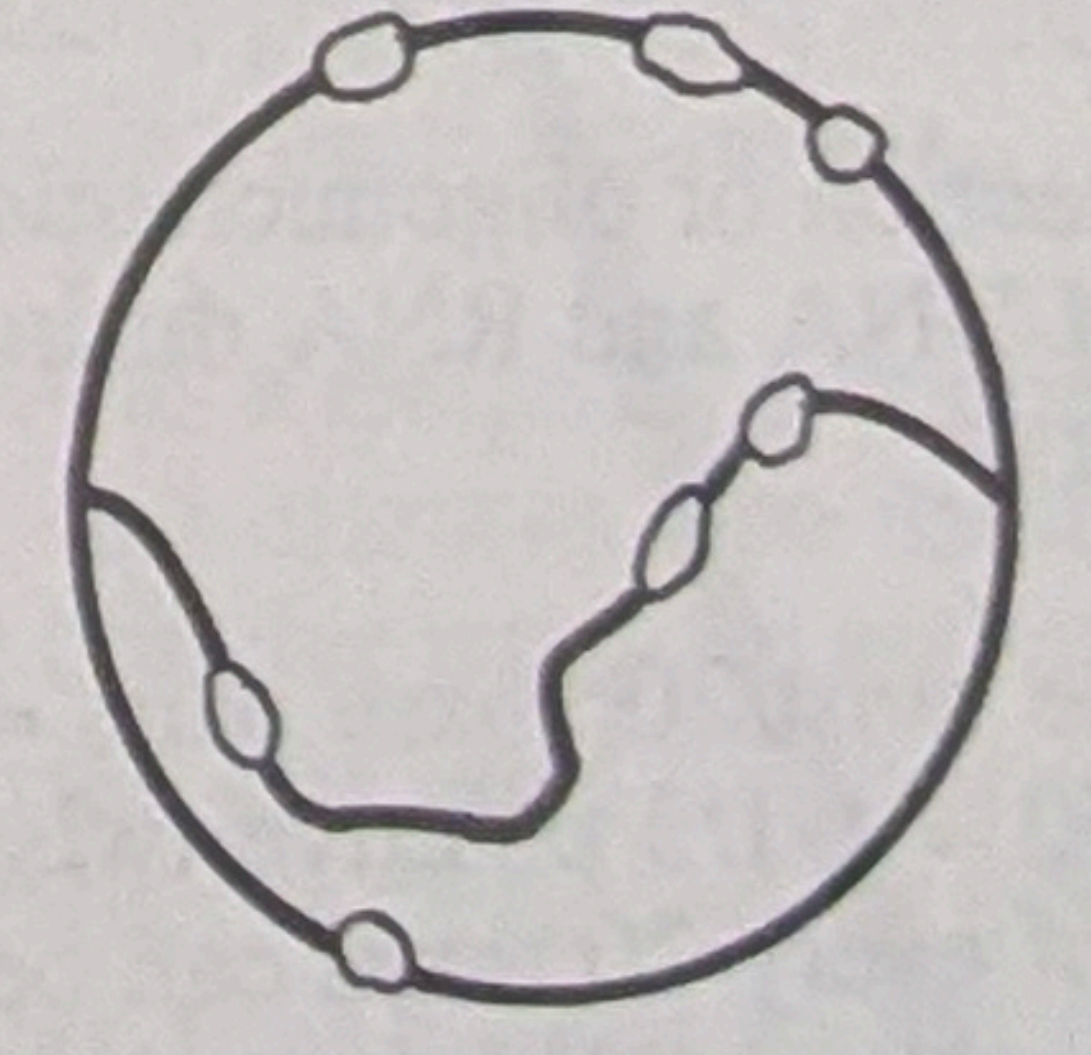
16 The molecular weight of a nucleotide pair is about 660, so the DNA molecule contains 38,000 nucleotide pairs. The replication rate of *E. coli* DNA is 50,000 nucleotide pairs per minute per replication fork. Thus, the time to replicate the phage DNA is (1/2)(38,000/50,000) = 0.38 min, or 22 sec.

17 Assume bidirectional replication and that both replication forks move at the same rate. Then the order is determined by the distance of each gene from the origin. Thus, the order is *AGBFCDE*.

18

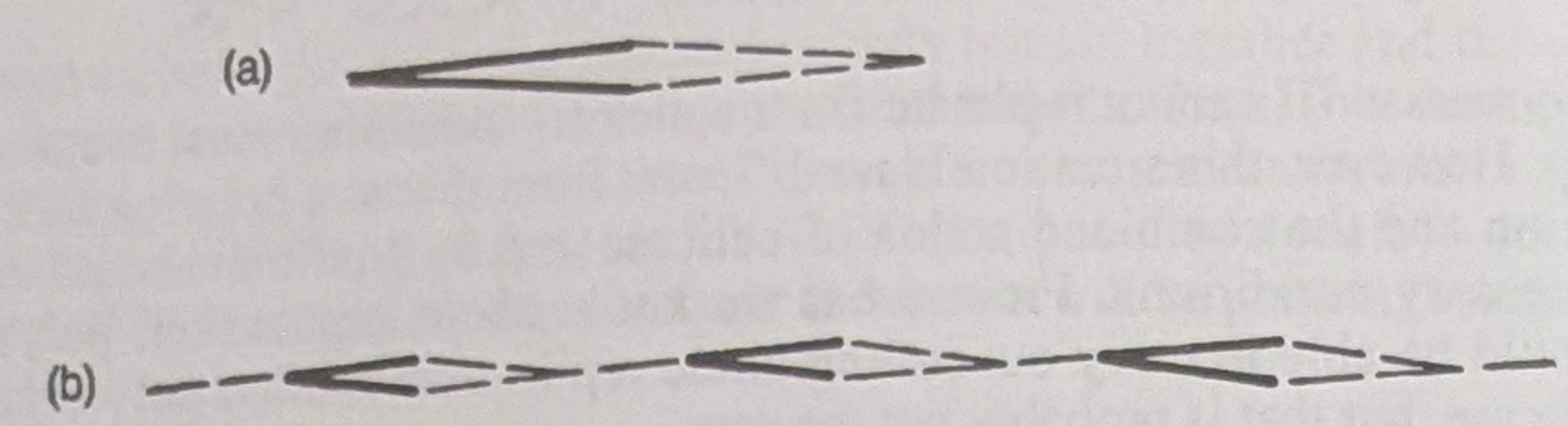


Unidirectional



Bidirectional

19



20 All modes require sites for initiation of all replication and sites for initiation of synthesis of precursor fragments. Unidirectional replication also needs a stop signal to prevent extension of the first precursor fragment; bidirectional replication does not need any other sites; single-stranded circular DNA needs a region to which SSB protein is not bound and at which priming can occur.

21 No, since in both cases the replication forks have passed the point which is halfway around the circle.

22 (a) They are joined normally by the combined action of polymerase I and DNA ligase, neither of which are temperature-sensitive. (b) Not many. A few fragments being extended by polymerase I might have a few ^{3}H-thymidines very near the 3' terminus.

23 The gap will be filled by a single piece. No fragments would be formed as long as one end of the gap had a 3'-OH group.

24 Bidirectionality of replication. If Okazaki fragments were produced in only branch of the fork, the branch would be different in each fork, so pulse-labeled material would hybridize to both parental strands.

25 Priming of precursor fragments occurs in the presence of rifampicin, which inhibits RNA polymerase and does not occur in a mutant with a defective *dnaG* gene, which encodes primase.

26 A collection of oligonucleotides of varied composition derived from degradation of DNA and RNA during purification. These oligonucleotides served as primers.

27 (a) One twist/200 base pairs = one twist per 20 turns of the helix. Similarly, 1 per 150 = 1-1/3 positive twists possible per 20 turns of the helix. Thus 2-1/3 turns of every 20 turns can be replicated or about 12%. (b) Only about 4% of mitochondrial DNA is replicated when the D-loop reaches its maximum size. Thus, DNA replication probably cannot proceed once the DNA is overwound.

28 Probably not, since there would be no twisting that would halt advance of the replication fork.

29 No, because there is no topological constraint to advancement of the replication fork.

30 The ligases form an active intermediate containing AMP by reacting with either NAD or ATP.

31 Since *E. coli* has two active replication forks and eukaryotic cells have thousands, one would expect animal cells to have several thousand times as many DNA polymerase molecules as a bacterial cell.

MORE CHALLENGING PROBLEMS

1 An *ung*$^-$ mutant fails to excise dU from DNA. If ^{3}H-labeled Okazaki fragments are isolated from an *ung*$^-$ mutant infected with phage λ and a hybridization is carried out with the individual complementary strands of λ DNA, will ^{3}H hybridize to only one or both complementary strands? Remember the mode of replication of λ DNA.

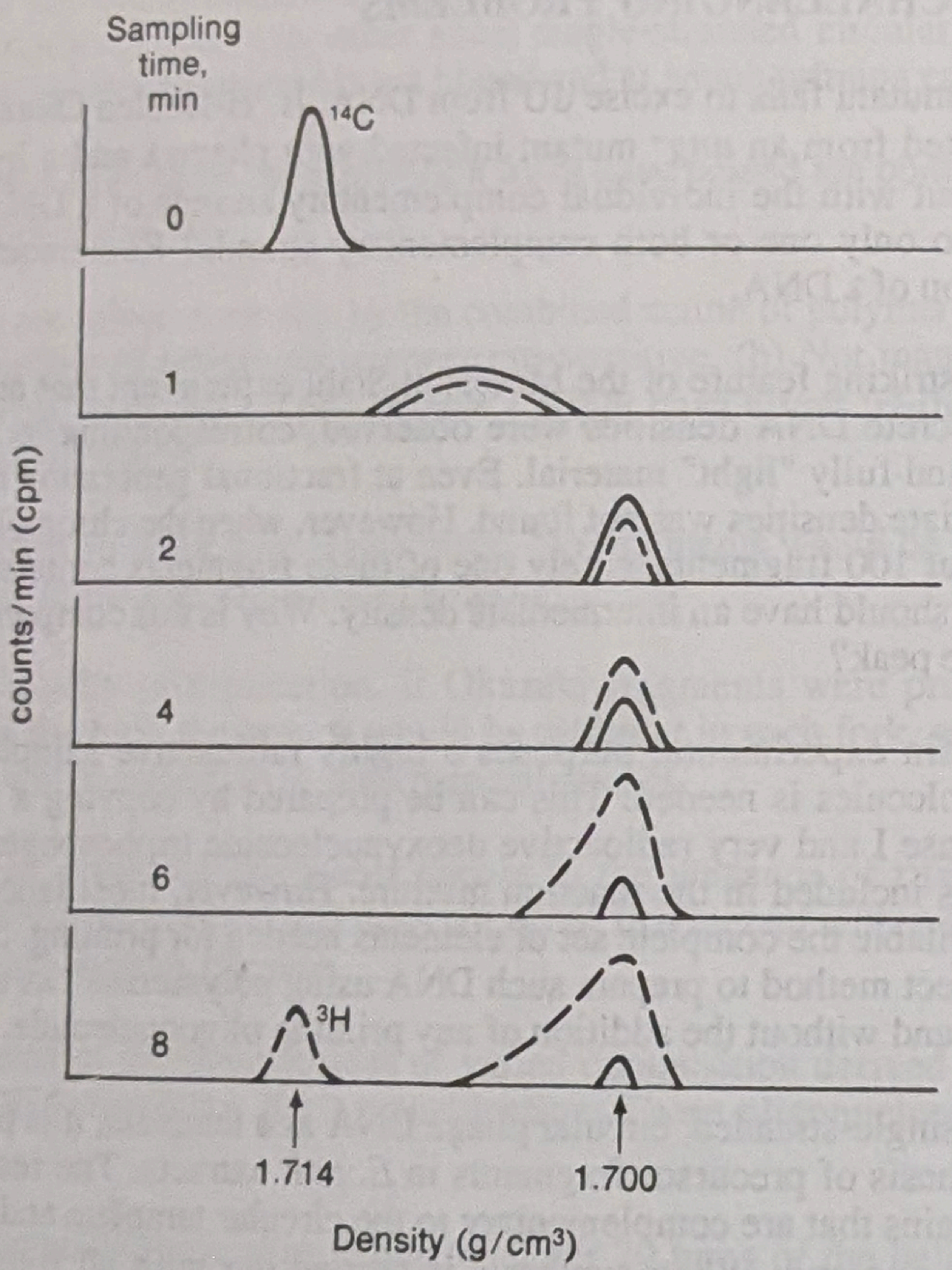
2 It was a striking feature of the Meselson-Stahl experiment that at all times only three discrete DNA densities were observed, corresponding to fully "heavy," hybrid, and fully "light" material. Even at fractional generation times, DNA of intermediate densities was not found. However, when the chromosome is broken into about 100 fragments, surely one of these fragments contains a replication fork and should have an intermediate density. Why is this component not seen as a discrete peak?

3 For certain experimental purposes a highly radioactive sample of particular DNA molecules is needed. This can be prepared by copying a template using polymerase I and very radioactive deoxynucleoside triphosphates, if a priming system is included in the reaction mixture. However, most laboratories do not have available the complete set of elements needed for priming. Suggest a much more direct method to prepare such DNA using polymerase I as the only protein element and without the addition of any priming oligonucleotide.

4 Using a single-stranded, circular phage DNA as a template, it is possible to carry out synthesis of precursor fragments in *E. coli* extracts. The reaction produces DNA chains that are complementary to the circular template and carry a tract of RNA on one end. When synthesis is carried out with all four α-^{32}P-labeled deoxynucleoside triphosphates, and the product is degraded by a procedure that yields 3'-ribonucleotides, radioactivity was found only in the adenine ribonucleotide. When only dCTP or dTTP was radioactive, no labeled ribonucleotides were detected. Use of labeled dGTP yielded four times as much

radioactive adenosine 3'-phosphate as when labeled dATP was used. (a) Is the RNA tract on the 3' or the 5' end of each new DNA strand? (b) What else can you conclude about the linkage of RNA to DNA? (c) What can be said about the nucleotide sequences on this DNA template at which polydeoxynucleotide synthesis initiates?

5 Consider a phage particle containing a small linear, single-stranded DNA molecule. Its replication mode is studied by centrifugation to equilibrium in CsCl. Its density in CsCl is 1.714 g/cm^3. Phage having ^{14}C-labeled DNA infect the bacterium in medium containing ^{3}H-thymidine. Samples are taken at various times, DNA is isolated, and each sample is centrifuged. The results shown in the figure below are obtained. (a) How does this phage replicate its DNA? (b) Do you think any progeny phage will be ^{14}C-labeled?



6 The single-stranded, circular DNA molecule of phage SC3 cannot be replicated by polymerase III unless RNA polymerase and four ribonucleoside triphosphates are added along with the four deoxyribonucleotides. The product is an open circular, double-stranded DNA containing a few ribonucleotides. In order to analyze the sequence of the initiating nucleotides, deoxyribonucleoside triphosphates labeled with ^{32}P in the α phosphate are used; the ribonucleoside triphosphates are unlabeled. The reaction product is treated with alkali and some radioactivity is released, but only into adenosine-3'-P (not the deoxy form). (a) What sequence information can be deduced from this result? (b) Next, α-^{32}P-dGTP is used with the other deoxynucleotides unlabeled, and the label is transferred only to adenosine-3'-P. What new sequence information is obtained? (c) If you use α-^{32}P-dGTP as the only labeled triphosphate and no ^{32}P appears in free ribonucleotides, what sequence information can be deduced? Where does the ^{32}P label end up in this case?

ANSWERS TO MORE CHALLENGING PROBLEMS

1 Only precursor fragments will be present but since λ replicates bidirectionally, they will be derived from both complementary strands. Thus, hybridization will occur to both strands.

2 Certainly the molecules having intermediate density exist. However, a problem exists in detecting these molecules. Since the chromosome is broken at random, the fraction of the DNA in the fork that is hybrid DNA can range from nearly zero to nearly one. Thus, the density of molecules containing the replication fork will range from heavy to hybrid. The molecules will be distributed uniformly in the intermediate density region rather than in a discrete peak. Furthermore, these molecules comprise so little of the material that they are not normally detected.

3 Nick the DNA lightly with an endonuclease and carry out nick translation. Note that you cannot start copying unnicked DNA from one end because there is no primer; you cannot even add a primer because in general you would not know the terminal base sequence.

4 (a) The 5' end. (b) The 3'-terminal ribonucleotide is A. (c) The nearest-neighbor deoxyribonucleotide to the adenine ribonucleotide is dG in 80% of the sequences, and dA in 20% of the sequences.

5 (a) First, the parent strand is copied to make double-stranded DNA. Hence, the density shifts to the lower value for native DNA. Between 2 and 4 minutes the amount of ^{3}H-labeled DNA increases with respect to ^{14}C-labeled-DNA but without a change in density. Thus, this must represent replication of the linear helix to form another double helix. This process continues at least until 6 minutes, as the amount of ^{3}H-double-stranded DNA has increased. Note that at 6 minutes some ^{3}H-DNA has a slightly higher density. This is due to displacement replication—that is, only one strand of the double-helix is being copied (the one

whose sequence is complementary to the ^{14}C-labeled strand), displacing a single strand. The displaced strands are released by 8 minutes. These are the progeny strands and presumably are packaged into phage particles. (b) Note that the ^{14}C strand remains at the density of double-stranded DNA. Thus, the initial ^{14}C-containing helix is used only as a template for synthesis of more double helices and no ^{14}C ever appears in progeny phage.

6 (a) The 3'-terminal base of the RNA primer is adenine. (b) The first DNA base added to the primer is dGMP, that is, 5'...rApdG...3'. (c) One can only conclude that dGMP is not the first deoxynucleotide added, showing that the primer does not have random ends. (d) The ^{32}P ends up in the undigested DNA.

CHAPTER 10

REPAIR

CHAPTER SUMMARY

DNA is subject to several types of damage. Base substitution can occur during replication, bases can be altered as a result of chemical instability of the bases themselves or of the *N*-glycosidic bond, and changes can occur as a result of environmental radiation and chemicals. The major changes that occur in DNA are (1) an incorrect base in one strand that cannot pair with the corresponding base in the complementary strand, (2) a missing base in one strand, (3) an altered base that is unable to form hydrogen bonds or that distorts the helix, single-strand breaks, double-strand breaks, and chemical cross links (covalent bonds between bases in complementary strands).

A variety of repair systems exist for correcting these classes of damage. Biological evidence for DNA repair comes from studies of survival curves, that is, graphs in which the surviving fraction of cells or viruses is plotted against the dose of the damaging agent. One type of experiment that indicates that a repair system is present is the increase in survival level following particular treatments. For example, if bacteria irradiated with ultraviolet (UV) light are exposed to visible light, the surviving fraction increases. This type of repair is called photoreactivation. Similarly, incubating UV-irradiated cells in a nonnutrient buffer for some time in the dark prior to placing the cells on agar increased survival, a phenomenon called dark repair. Another type of experiment utilizes mutants that are radiosensitive. For example, *uvrA*$^-$ and *recA*$^-$ mutants of *E. coli* are more easily killed by UV than are wild-type cells, indicating that the *uvrA* and *recA* genes are involved in repair processes. A *uvrA*$^-$*recA*$^-$ double mutation confers greater sensitivity than either of the individual mutations, which indicates that there are two distinct repair systems.

The major lesion produced by UV light is the pyrimidine dimer, of which the thymine dimer has been the most carefully studied. A pyrimidine dimer is produced by forming two bonds between the rings of two adjacent pyrimidines in the same DNA strand. The two bonds form a four-membered ring, so the dimers are sometimes called cyclobutane-type dimers. Thymine dimers block DNA replication and their effects must be dealt with if an irradiated cell is to survive. There are four major mechanisms for repair of thymine dimers: photoreactivation, excision repair, recombination repair, and SOS repair. Photoreactivation is a direct cleavage of the pyrimidine dimers. It is an enzymatic process in which the enzyme, photolyase, is somehow activated by blue and near-ultraviolet light. Excision repair is a multistep process in which certain types of altered bases are excised by the combined action of a glycosylase, an incision endonuclease and an exonuclease, followed by repolymerization (by polymerase I) of the gap produced by the excision. In *E. coli* incision is carried out by the Uvr enzyme. The final step is sealing by DNA ligase. Recombination repair is an exchange process occurring in replicating molecules; gaps in one daughter strand produced by aberrant replication across damaged sites (initiation at a random site, possibly without an RNA primer) are filled in by nondefective segments from the parental strand of the other branch of the newly replicated DNA. Recombination repair depends on an active RecA product.

The SOS repair system is both inducible by UV light and is error-prone. Induction is an indirect result of the presence of UV-irradiated DNA, which causes increased synthesis of the RecA protein, a phenomenon described in Chapter 15. When DNA polymerase III encounters a pyrimidine dimer, the distortion activates the editing function, causing the polymerase to remain at the site, alternately inserting and excising an adenine nucleotide. However, damaged DNA binds the RecA product, which contacts polymerase III, alters its structure, and causes it to lose its editing function. In that state, pol III can proceed onward. After leaving the dimer site, pol III, no longer in contact with the RecA protein, regains its normal editing activity. However, because of the lack of editing at the dimer site, rare incorporation errors at the site are not repaired. Hence, SOS repair synthesis frequently gives rise to mutations. Error-prone repair is the major cause of mutagenesis by ultraviolet radiation and by alkylating agents.

Photoreactivation, excision repair, and recombination repair do not lead to mutation.

BOLD TERMS

alkylating agents, bypass system, dark repair, daughter-strand gap repair, depurination, excision repair, incision, liquid-holding recovery, photolyase, photoreactivation, postdimer initiation, PR enzyme, recombination repair, sister-strand exchange, SOS repair, survival curve, transdimer synthesis, UV-reactivation, W reactivation

ADDITIONAL INFORMATION

Four systems for repair of pyrimidine dimers—photoreactivation, excision repair, recombination repair, and SOS repair—have been described in the text. The features of each will be summarized here. These systems differ both in the source of energy used for repair—blue and near-ultraviolet light in the case of photoreactivation and ATP molecules for the other three—and in the mechanisms used to form functional daughter DNA molecules. Excision repair and photoreactivation both repair the template, whereas recombination repair forms a new template. SOS repair instead ignores the damage and forms uninterrupted daughter strands, despite the presence of damaged segments of the template. Of the four, only SOS repair is mutagenic.

Cross-links have been experimentally introduced into bacterial DNA in two ways—by addition of antibiotics such as mitomycin C, and by photochemical mechanisms in which certain intercalating molecules such as 8-methoxypsoralen are added to the suspending buffer after which cells are irradiated with near-UV light. When survival of bacteria treated with these agents is plotted against time of exposure to a particular concentration of mitomycin C or to a dose of light following psoralen treatment, it is found that Uvr$^-$ and RecA$^-$ mutants are much more sensitive than wild-type *E. coli*, indicating that wild-type *E. coli* possesses a system for repairing cross-links, which utilizes the *uvrA* and *recA* genes. At the present time, several hypotheses have been put forth for the mechanism of repair of cross-links, but none has supporting evidence. Repair of cross-links must also occur in mammalian cells, because cultured cells obtained from patients with Fanconi's anemia (a bone marrow disease) are much more sensitive to psoralen-induced killing than are cells obtained from a normal human.

Efficient systems also exist for repair of the damage produced by ionizing radiation, namely, single-strand and double-strand breaks and the production of the altered base 5,6-dihydroxydihydrothymine (DHDT). Repair of single-strand breaks introduced by ionizing radiation has been carefully studied. A few of the breaks have free 3'-OH and 5'-P groups without any other alterations and are repaired by DNA ligase alone. However, mutant cells lacking polymerase I (*polA*$^-$ mutants) are more sensitive to x-irradiation, which suggests that polymerase I-mediated repair replication is needed. Cells mutant only in the 5'-3'-exonuclease activity of polymerase I are also defective in this repair; the reason is that base damage invariably has occurred near each single-strand break. The repair of base damage, including that which is not adjacent to breaks, has been examined by studying the removal of DHDT. This process is not as well understood as thymine-dimer excision but in both *E. coli* and in mammalian cells an endonuclease activity has been found that makes cuts near DHDT (and other altered bases). This enzyme is not the thymine dimer-incision nuclease inasmuch as *uvrA*$^-$ cells are not sensitive to x rays and the endonuclease activity is not reduced in a *uvrA*$^-$ cell. The incision event is followed by excision by the 5'-3' exonuclease of polymerase I, after which repair replication mediated by polymerase I and DNA ligase occurs. This process is probably required, at least in part, for repair of single-strand breaks accompanied by base damage. A DHDT-*N*-glycosylase is also present in *E. coli*. Cells that are *recA*$^-$ are also extraordinarily sensitive to x rays; thus, either recombination repair or some product of the SOS system plays a role in the repair of x-ray damage. Evidence from cultured cells obtained from diseased humans shows that repair of x-ray damage occurs in mammalian cells. In this case the cells are obtained from an inherited retinoblastoma, a cancer of the retina; these cells are hypersensitive to x rays and unable to repair single-strand breaks efficiently.

There is little solid evidence for direct repair of double-strand breaks. Since there is no obvious way to bring together two freely moving ends, the few supporting observations have been viewed with suspicion. Since it is believed that all DNA molecules, including those of bacteria, are associated with structural proteins, it is questionable whether there really are two freely moving ends at the site of a potential break. Thus, the view is that when repair is observed, the two ends have been held adjacent to one another by these proteins. Further experiments are needed to resolve this question. In yeast there is good evidence for a complex mechanism for repair of double-strand breaks that utilizes recombination, as is described in Chapter 21 of the text.

DRILL QUESTIONS

1 Name four ways that the base sequence of DNA can be changed.

2 Name the two most common changes in DNA that occur spontaneously without attack by external agents.

3 Are thymine dimers formed between adjacent thymines in the same strand or between opposite thymines in complementary strands?

4 Do two single-strand breaks have to be directly opposite one another to constitute a double-strand break?

5 If UV-irradiated bacteria are incubated in a buffer prior to plating, will the number of colonies formed be increased or decreased by the treatment?

6 If UV-irradiated phage are incubated in a buffer or exposed to light prior to plating, how will the efficiency of plaque formation change?

7 Which repair system cleaves thymine dimers?

8 What is the purpose of light in the functioning of the photoreactivating enzyme?

9 Which repair system cleaves *N*-glycosidic bonds?

10 Which repair system cleaves phosphodiester bonds?

11 Which enzymes are required for excision repair in *E. coli*?

12 What is the difference between incision and excision?

13 Name two features of SOS repair that distinguish it from other repair systems.

14 When the effect of pyrimidine dimers is bypassed by the SOS repair system, are the mutations produced normally at the site of the dimer?

15 What genes in *E. coli* are necessary for SOS repair?

16 Can an adenine nucleotide still base-pair with a thymine in a thymine dimer?

17 Why does idling occur at an unrepaired dimer site?

ANSWERS TO DRILL QUESTIONS

1 Replication errors, repair errors, loss of bases due to chemical instability, attack by environmental agents.

2 Spontaneous deamination of cytosine and depurination.

3 In the same strand.

4 No. They can be separated by a few base pairs.

5 Increased.

6 It will be unaffected since phage particles do not contain any repair enzymes.

7 Photoreactivation.

8 It provides activation energy for the dimer-cleavage reaction.

9 A base-*N*-glycosylase.

10 Excision repair.

11 The Uvr endonuclease, polymerase I, and DNA ligase.

12 Incision refers to the production of one single-strand break, usually near a nucleotide that is to be removed. Excision refers to the second break and subsequent removal of a nucleotide or a set of adjacent nucleotides.

13 It is inducible and it is error-prone.

14 Yes.

15 *RecA* and the *umu* genes.

16 Yes. The hydrogen-bonding groups are not altered in the dimer.

17 Even though adenine can still pair with a thymine in the dimer, the base pair is weakened by the distortion of the helix, and polymerase I response to the weak pairing by activating its editing function.

PROBLEMS

1 Do you think that photolyase cleaves all thymine dimers in an ultraviolet-irradiated cell?

2 Would you expect photoreactivation to occur with equal efficiency over a wide range of wavelengths or will it exhibit a spectrum with maximum effectiveness in a small range of wavelengths?

3 A cell possesses a repair system capable of removing half of the damage produced by some agent. Which of the following statements would be true of a survival curve: (1) a dose yielding *x* percent survival when repair does not occur would yield 2*x* percent survival if repair occurs; (2) the dose required for a particular percentage of survival in the absence of repair is doubled when repair occurs.

4 Suppose 30 years ago you had observed that DNA replication in bacteria is inhibited by radiation from a germicidal lamp that emits ultraviolet light. Clearly the damage could result from direct absorption of the ultraviolet light by the DNA, but it was necessary to consider the possibility that some other molecule absorbed the light and afterwards damaged the DNA. Without using modern physical and chemical techniques what simple spectral measurement might you have done to investigate this question?

5 The ability of ultraviolet-irradiated T4 phage to form plaques is almost the same on both Uvr$^+$ and Uvr$^-$ bacteria, with the survival curve on the Uvr$^-$ bacteria being only slightly steeper than the one obtained with Uvr$^+$ bacteria. (a) How might you explain this fact? (b) A T4 mutant is UV-irradiated and plated on Uvr$^+$ and Uvr$^-$ bacteria. The survival curves are much steeper than with wild-type phage and considerably steeper with Uvr$^-$bacteria than with Uvr$^+$ bacteria. Suggest a defect for the mutant.

6 A bacterial repair system called X removes thymine dimers. You have in your bacterial collection the wild-type (X$^+$) and an X$^-$mutant. Phage λ, when ultraviolet-irradiated and then plated, gives a larger number of plaques on X$^+$ than on X$^-$ bacteria. It has been proposed on the basis of survival curve analysis that the X enzyme is inducible. To test this proposal, ultraviolet-irradiated λ phage are adsorbed to both X$^+$ and X$^-$ bacteria in the presence of the antibiotic chloramphenicol (which inhibits protein synthesis). No thymine dimers are removed in the X$^-$ cell and 50% are removed in the X$^+$ cell. In the absence of chloramphenicol, the same results were obtained. (a) Is X an inducible system? (b) Suppose 5% of the thymine dimers are removed in the presence of chloramphenicol and 50% in its absence; how would your conclusion be changed?

7 A Uvr$^-$ bacterial culture that has been grown for many generations in medium containing [^{3}H]thymidine is transferred to nonradioactive medium. The culture is UV-irradiated; since the cells are Uvr$^-$, no radioactive material appears in the medium. A phage sample is then used to infect the irradiated culture, and several minutes later radioactive thymine dimers are found in the medium. The phage does not cause degradation of bacterial DNA from unirradiated bacteria. Explain the appearance of the dimers.

8 It is generally observed that if the UV survival curves of two different phages are compared, the phage with the larger DNA molecule has the steeper curve. (a) Explain. (b) You have isolated a new phage whose DNA is very large but which has the radiosensitivity of a phage with a small DNA. Suggest a possible explanation.

9 The radiosensitivity (measured by colony formation) of bacteria obtained from a log-phase culture and from late stationary phase is measured. Both give curves with the general shape shown in Figure 10-5 of the text, but the cells in log phase are more resistant to the radiation. Suggest two possible explanations, one of which includes DNA repair and one of which does not.

ANSWERS TO PROBLEMS

1 It seems unlikely, since light-induced repair is never complete. An alternative is that photolyase only repairs thymine dimers and does cleave all of them, and that half of the killing is due to other pyrimidine dimers. The fact is that all thymine dimers are not cleaved. Perhaps cleavage can only occur when the dimers are adjacent to certain types of sequences.

2 Like all light-absorbing phenomena, there is a photoreceptor that absorbs in a particular range of wavelengths. Thus, we should expect a wavelength-response curve.

3 Statement 2.

4 Use a set of filters that enable you to irradiate the cells with a particular range of wavelengths. Measure the effectiveness of killing (e.g., the dose yielding a

particular survival value) as a function of wavelength. The curve showing this function is called an action spectrum. If DNA is the primary absorber, the action spectrum should match the absorption spectrum of DNA.

5 (a) Either the damage to T4 DNA is, for some reason, barely repairable by the Uvr system or T4 possesses its own repair system. (b) The mutant probably lacks a T4-encoded repair system.

6 (a) X is not inducible, because the enzymes were present before protein synthesis was blocked by chloramphenicol. (b) X would probably be considered inducible. The residual 5% could be due to either a second noninducible system, which can excise thymine dimers, or to a small amount of synthesis of X proteins when chloramphenicol is present.

7 The phage apparently makes its own incision enzyme.

8 (a) The larger the DNA molecule, the greater is the probability that the DNA will receive a lethal alteration. (b) The phage may possess its own repair system.

9 A simple non-repair explanation is that in a log-phase culture many cells are paired and both cells have to be killed to prevent colony formation. An alternative is that the repair systems are more active in rapidly multiplying cells.

MORE CHALLENGING PROBLEMS

1 Cells having a very low concentration of either polymerase I or DNA ligase are somewhat more sensitive to ultraviolet light than wild-type cells but not as sensitive as Uvr$^-$ mutants. How is this fact compatible with the fact that both polymerase I and DNA ligase are needed in excision repair?

2 If wild-type λ phage are plated on *E. coli*, the resulting plaques have a turbid center. About 0.01% of the plaques are completely clear because they harbor a mutation in one of three different genes that determine plaque turbidity. (a) The phage are UV-irradiated and then allowed to go through a cycle of infection in UV-irradiated wild-type bacteria. The progeny phage are plated and the frequency of clear plaques is observed to be 0.2%. Why? (b) A radiosensitive *E. coli* mutant is also irradiated and infected with the original irradiated phage. The yield of phage is much lower than with wild-type bacteria, and among the progeny phage, 0.01% produce clear plaques. Explain.

3 A Uvr$^-$ bacterial culture that has been grown for many generations in medium containing $[^3H]$thymidine is transferred to nonradioactive medium. The culture is then UV-irradiated, but since the cells are Uvr$^-$, no radioactive material appears in the medium. Puromycin, an antibiotic inhibitor of protein synthesis, is added to the culture, after which the culture is infected with a phage sample. The puromycin prevents phage production and all phage-induced protein synthesis. Nonetheless, several minutes after phage infection radioactive thymine dimers are found in the medium. The phage does not cause degradation of bacterial DNA from unirradiated bacteria. Explain the appearance of the dimers.

ANSWERS TO MORE CHALLENGING PROBLEMS

1 Pol I and ligase are essential for normal DNA replication. Therefore, the mutants, which are viable, must have a residual activity of these enzymes and this activity must be sufficient to carry out some excision repair.

2 (a) Irradiation of the bacteria has activated the SOS repair system, which is error-prone and, hence, in repairing the dimers in the irradiated phage has produced an increased frequency of mutant progeny phage. (b) The bacterial mutant contains a mutation in one of the genes needed for SOS repair. Hence, a greater fraction of the phage are not repaired and none are repaired by an error-prone system.

3 The phage clearly could not have synthesized either the incision enzyme or a thymine dimer glycosylase, so the enzyme must have existed before the infection. It could not have come from the bacterium, so it must have been present in the phage head and injected with the DNA.

CHAPTER 11

MUTAGENESIS, MUTATIONS, AND MUTANTS

CHAPTER SUMMARY

A mutation is a change in the base sequence of the DNA of an organism. The change may involve a single base pair (a point mutation), or several. The wild-type base pair may be replaced by another base pair (a base substitution). If the substitution is a purine for a purine and a pyrimidine for a pyrimidine, the mutation is called a transition; if a purine is replaced by a pyrimidine, or vice versa, the mutation is called a transversion. Transitions occur much more frequently than transversions. Some mutations involve loss of one or more base pairs (a deletion), addition of base pairs, inversions of base sequences, or relocation of base sequences (translocations).

A point mutation in a coding sequence may or may not cause an alteration in the amino acid sequence of a protein; when an amino acid change occurs (a missense mutation), it may or may not affect the activity of the protein. If a change in activity occurs, the phenotype of the organism is said to be altered. Inversions, translocations, base additions, and deletions invariably cause phenotypic changes unless they occur in noncoding regions of DNA. If there is no phenotypic change in an organism, the mutation is said to be silent. If the phenotypic change is so small that the effect on the organism is insignificant, the mutation is leaky. In some cases a point mutation leads to a phenotypic change only in particular conditions. Such mutations are conditional mutations. One example is a temperature-sensitive mutation, in which the phenotypic change exists only above or below a critical temperature. Most temperature-sensitive mutations are heat-sensitive. Other conditional mutations are only expressed when other mutations are not present; these are usually suppressor mutations. One class of point mutation, the chain-termination or nonsense mutation, causes premature termination of synthesis of a protein. These are frequently conditional mutations. An organism in which a mutation is expressed is a mutant.

Mutations are produced by mutagenesis, which may be spontaneous or induced. Spontaneous mutations arise primarily by replication errors. Induced mutations are a result of exposure of the organism to precursors that substitute

for normal bases, to radiation, or to chemicals that react with DNA bases.

Spontaneous mutations may occur, though at very low frequency, by misincorporation of a base, but these errors are usually corrected either by the editing system or by mismatch repair. Some spontaneous mutations result from deamination of 5-methylcytosine, deletion or addition of a base during replication, intramolecular recombination between directly repeated sequences flanking a unique sequence, or replication across cruciforms formed by inverted repeats. Mutations can be induced in a variety of ways. Ultraviolet radiation induces mutations by activating the error-prone SOS repair system. Ionizing radiation causes chemical alterations of bases in DNA. Nitrous acid converts a normal DNA base to one with different hydrogen-bonding specificity, thereby leading to transitions. Alkylating agents and low pH cause production of depurinated sites; when these are copied, adenine is frequently inserted in the daughter DNA strand, independently of the identity of the base removed. Base analogues have two tautomeric forms with different hydrogen-bonding specificity; they are usually incorporated in one form and then spontaneously assume another form; if this occurs when the base is serving as a template for further synthesis, a mutation will result. 5-Bromouracil and 2-aminopurine are commonly used base analogues. 5-Bromouracil also causes imbalances in internal metabolic pools of nucleotides and thereby increases the frequency of spontaneous misincorporation. Acridines intercalate between base pairs and cause base additions and base deletions during replication. Some organisms have an aberrantly high frequency of mutation. This can be traced to the presence of mutators, which are often mutant forms of genes involved in DNA replication or repair. For example, *dam* mutants are unable to distinguish parental and daughter strands during mismatch repair, so correct bases are changed to incorrect ones as often as incorrect bases are corrected. A DNA polymerase with reduced editing activity is another example of a mutator.

Site-specific mutagenesis refers to selective alteration of a particular base pair in a specific base sequence. It is an important tool in genetic engineering. A variety of procedures are in current use.

Reversion is a process by which a normal phenotype is restored to a mutant. It can result from restoration of the wild-type base sequence, but usually it occurs by spontaneous production of mutations at other sites. An example of a second-site reverse mutation is one that restores the three-dimensional shape of a protein by allowing two amino acids at particular positions to interact. For example, the normal interaction of amino acids A and B might have been eliminated by an alteration of A to X, which cannot interact with B; a second-site mutation converting B to Y, which can interact with X, is an example of a reverse mutation. Such a reversion event is an intragenic reversion. Some reversions are mutations in other genes (intergenic reversion); an example is the production of a compensating mutation in the binding site of a polypeptide that is a component of a multisubunit protein. Intergenic reversions that suppress the effect of a large number of mutations are called suppressors, they are described in Chapter 13. Measurement of reversion frequency is a useful method for screening for mutagens and since most carcinogens are also mutagens, for carcinogens also. One test, the Ames test, is widely used; it incorporates an extract of rat liver into growth media and enables one to detect substances that only become mutagenic after conversion in the liver.

Some mutations occur in the target sequence for a restriction enzyme. These are called restriction polymorphisms. They are exceedingly useful in human genetics as a means of detecting a mutant allele in a heterozygous individual. Carries of the sickle-cell and Huntington trait have been identified by screening for restriction polymorphisms.

BOLD TERMS

alkylating agent, amber mutation, Ames test, base deletion, base insertion, back mutation, base substitution, chain termination mutation, depurination, frameshift mutations, hot spots, induced mutagenesis, intercalation, leaky mutation, microsomal fraction, missense mutation, mutagen, mutagenesis, mutant, mutation, mutator genes, nonsense mutation, ochre, opal, phenotype, point mutation, restriction polymorphism, reverse mutation, reversion, revertant, second-site revertant, silent mutation, spontaneous mutagenesis, spontaneous mutation, suppressor mutation, temperature-sensitive, transition, transversion, Ts mutation

ADDITIONAL INFORMATION

Mutants are usually isolated from large populations of organisms and hence some means is needed to distinguish a mutant and a wild-type organism. Furthermore, the mutants may constitute an extremely small fraction of the total population, so that prior to searching for a mutant it is necessary to increase the mutant fraction. Mutants can be detected in several ways. The following two procedures are used to detect bacterial mutants that are either resistant to an antibiotic or unable to utilize a particular sugar:

1. An antibiotic-resistant (Ant-r) mutant is found simply by plating a large number of bacteria on a medium containing the antibiotic. Only an Ant-r cell can form a colony. This method applies to the isolation of mutants resistant to any chemical compound that can be incorporated into agar.
2. Sugar-utilization mutants (we consider lactose as an example) are isolated by means of color indicator solid medium. A common medium, called EMB agar, consists of a complex broth containing a very high concentration of all amino acids, lactose, and the two pH-sensitive dyes, eosin and methylene blue. Lac$^+$ cells preferentially use the lactose as a carbon source and, in metabolizing lactose oxidatively, excrete H$^+$ ions. Thus, the pH of the medium surrounding the cells decreases, causing the dyes that permeate the colony to become a deep purple. Lac$^-$ cells cannot utilize the lactose but can use several amino acids as a carbon source. NH_3 is released, and the pH increases locally around the cells, decolorizing the dyes and producing a white colony. Thus, on EMB agar Lac$^+$ colonies are purple and Lac$^-$ colonies are white. These color differences occur for all metabolizable sugars and for many other carbon sources as well.

The sugar-utilization mutants just described can grow in a broth lacking the sugar, if another carbon source is provided. However, if we seek a leucine-requiring (Leu$^-$) mutant, we cannot use a medium lacking leucine, because the desired mutant will not grow. The following technique is useful, in principle. If a mixture of 20 Leu$^+$ cells and 5 Leu$^-$ cells were plated on a minimal agar supplemented with leucine (we call this the master plate), 25 colonies would form. If a second petri dish were prepared containing minimal agar lacking leucine (a test plate) and, by using 25 sterile toothpicks, a tiny portion of each Leu$^+$ colony was transferred from the master plate to the test plate, only 20 of the transfers would form a colony. If one kept track of the initial colonies transferred, the five Leu$^-$ colonies on the master plate would be identified. Such a technique would be extremely tedious if the fraction of cells that was mutant was very small. The replica-plating procedure enables one to transfer all colonies from the master plate to test plates in a single step, in the following way. A circular piece of sterile velvet is pressed onto the leucine-containing master plate and then picked up. A portion of each colony is retained in the hairs of the cloth. The velvet is then pressed onto the surface of two fresh plates, one with leucine and one without leucine, and then colonies are allowed to grow. Leu$^+$ colonies form on both plates, but the Leu$^-$ cells fail to grow on the leucine-deficient agar, and the Leu$^-$ colonies are identified as those which grow only on the agar containing leucine. Typically one would not be seeking 5 mutants in 25 but 1 in 10^4; by replica-plating, 50 plates having 200 colonies on each plate would suffice, and this is not an unmanageable number.

Phage mutants are detected in two ways. "Plaque-morphology" mutants are found because they look different from wild-type plaques. For example, some mutants make plaques that are turbid, or clear, or have a halo, or are a combination of these. Conditional mutants, that is, mutants that are either temperature-sensitive or host-dependent (forming plaques on one host but not on another), are commonly used and can also be found by replica-plating. With a temperature-sensitive mutant, plaques are formed at 30 C and then transferred to two plates already seeded with bacteria, and incubated at either 30 C or 42 C. A temperature-sensitive mutant forms a plaque on the 30 C plate but not on the 42 C plate.

If a population of bacteria contains mutants at a frequency of 1 per 10^7, a not unreasonable number, 50,000 plates would have to be replicated in order to find one Leu$^-$ mutant. Thus, mutant enrichment procedures, which favor the growth of the mutant over the wild type, are used. One such procedure is the penicillin selection technique. This method, which is applicable only to bacteria, is based upon the fact that penicillin kills bacteria by interfering with the synthesis of the bacterial cell wall. When penicillin is added to a culture of growing bacteria, cell wall synthesis stops but growth (enlargement) of the cell continues until the cell bursts. A Leu⁻ cell in a growth medium lacking leucine cannot synthesize protein and thus cannot enlarge, so it is not killed by penicillin. Thus, growth of a culture of Leu⁺ cells containing rare Leu⁻ mutants will be considerably enriched for the mutant simply by growth of the culture in medium lacking leucine and containing penicillin. Enrichment is not 100% because the bursting cells release nutrients that would cause the mutant cells to grow and hence be killed by the penicillin.

Reversion studies are useful to characterize mutations. For example, double mutants and deletions can be distinguished from point mutants in that they are not revertable at a detectable frequency. The kind of base change can also on occasion be identified by using mutagen-stimulated reversion. For example, the mutagens 5-bromouracil (BU) and 2-aminopurine both induce a change from an AT pair to a GC pair. If a cell containing a 5-bromouracil-induced mutation is mutagenized by growth in the presence of 2-aminopurine, reversion is stimulated because second-site base substitutions occur. The reversion frequency is even greater if the mutagen is hydroxylamine, which causes GC-to-AT transitions, because the hydroxylamine not only causes second-site reversions but also sometimes restores the original AT pair by altering the mutant GC pair. Similarly, hydroxylamine-induced mutants are stimulated to revert by hydroxylamine but more so by growth in medium containing 5-bromouracil or 2-aminopurine. In summary, base substitution mutations can be reverted by any base-substituting mutagen but reversion frequency is greater if there is a significant potential for restoring the original base sequence. Growth in medium containing acridines, which cause frameshift mutations, would not generally stimulate reversion of base substitution mutations. On the other hand, frameshift mutations are stimulated to revert by acridines but not by base-substituting mutagens. Whether a mutation can be reverted by a base substitution mutagen or by an acridine is a useful criterion for identifying a base substitution or a frameshift mutation, respectively.

DRILL QUESTIONS

1 Is a base-pair change always a mutation?

2 Does a base-pair change necessarily change the phenotype?

3 Which of the following base-pair changes are transitions and which are transversions: (1) AT→ TA, (2) AT→ GC, (3) GC→TA?

4 Do base analogues produce transitions or transversions?

5 Do most chemicals that alter the chemical structure of a base cause transitions or transversions?

6 What base is most often inserted when a depurinated site is in the template?

7 Distinguish a missense and a nonsense mutation.

8 What makes a particular mutation temperature-sensitive?

9 In the notation for bacterial mutations how are phenotypes and genotypes distinguished?

10 In gene symbols what do capital-letter suffixes, as in *proA*, *proB*, etc. mean?

11 Distinguish mutation and mutant.

12 What type of mutation is induced by acridines?

13 What type of mutation is induced by both nitrous acid and by 5-bromouracil?

14 How can tautomerization cause mutation?

15 By what mechanisms does 5-bromouracil induce mutations?

16 Does a frameshift always cause a phenotypic change?

17 A deletion occurs that eliminates a single amino acid in a protein. How many base pairs were deleted?

18 Mutations in what system increase the spontaneous mutation frequency?

19 What base is responsible for hot spots for spontaneous point mutations?

20 A mutant is isolated that cannot be reverted. What biochemical type(s) of mutation might it carry?

21 What class of point mutation is not reverted by second-site intragenic reversion?

22 In general what can be said about the location of a revertant of a frameshift mutation?

23 Why is a liver microsomal fraction included in the Ames test for mutagens?

24 Will every mutation generate a restriction polymorphism?

ANSWERS TO DRILL QUESTIONS

1 Yes, because, strictly speaking, that is the definition of a mutation.

2 No. The change does not always cause an amino acid replacement, and an amino acid change does not always alter protein function.

3 Transition, 2; transversion, 1, 3.

4 Transitions.

5 Transitions, unless the alteration leads to breakage of the *N*-glycosidic bond or activates the SOS repair system.

6 Adenine.

7 A missense mutation results in an amino acid substitution. A nonsense mutation results in premature termination of a protein.

8 An amino acid substitution in the protein product causes weakening of the structure such that a shape change occur in the protein above or below a critical temperature.

9 Phenotype uses a capitalized, roman symbol; genotype uses a lower-case italicized symbol.

10 The suffixes distinguish genes that are in a single biochemical pathway or process, in this case, proline biosynthesis.

11 Strictly speaking, a mutant is an organism carrying a mutation (a change in nucleotide sequence). However, the term mutant is usually used only if the mutation is expressed.

12 A frameshift.

13 A transition.

14 A tautomer can cause mutations if it has two forms in equilibrium with another and each form has different base-pairing properties. A tautomeric molecule can be incorporated in an aberrant form and then switch to a form that cannot pair with the base in the DNA. Alternatively, when in the template strand, it can temporarily switch its base-pairing properties and allow incorporation of a base that would not have been added to the end of the growing chain if the switch had not occurred.

15 It tautomerizes, which allows base substitutions, and it affects nucleotide pools, which increases incorporation errors.

16 Yes, because the entire frame of reading of the bases is changed, which results in alterations of most amino acids downstream from the mutation. Only two types of base addition or deletion would not cause a phenotypic change: one that occurs in the space between genes and one that occurs very near the end of a gene.

17 Three. Otherwise, there would have been a frameshift and all downstream bases would have been changed.

18 Editing and mismatch repair.

19 5-Methylcytosine.

20 A frameshift or a deletion.

21 Chain termination mutations, since the reversion event would have to eliminate the chain termination site. Of course, since a chain termination codon contains three bases, it is possible that a second-site reversion could occur in one of the two positions of the chain termination codon other than that which had originally been altered. However, this would certainly produce an amino acid change at the mutated site, so reversion would be unlikely.

22 Typically, it is a second-site mutation, but the site must be fairly near the original mutation; otherwise, many amino acids would be changed.

23 Many substances are not mutagenic by themselves but are converted to mutagens by enzymes in the liver.

24 No, because the base change will not always be in a target sequence for a restriction enzyme; in fact, this will usually not be the case.

PROBLEMS

1 Consider a DNA molecule with the structure shown below. (a) If during replication, the T at position 2 is read as an A, what are the results of the first and second rounds of replication? (b) If the replication apparatus copies the AT pair at 2, and then copies the AT at 5, skipping 3 and 4, what are the results of the first and second rounds of replication? Only bases 1 through 6 are shown.

```
A T G C A T T A T G C C
T A C G T A A T A C G G
          1 2 3 4 5 6
```

2 One variety of a temperature-sensitive mutation is the cold-sensitive (Cs) mutation, which has a mutant phenotype below a particular temperature. Bacteria containing the mutation *ess2*(Ts) can form colonies at 32 C, but not at 37 C and 42 C, whereas those containing the mutation *ess5*(Cs) form colonies at 42 C, but not at 32 C and 37 C. What would be the phenotype of an *ess2*(Ts)*ess5*(Cs) double mutant?

3 Often dyes are incorporated into agar in order to determine whether a bacterium can utilize a particular sugar as a carbon source. For instance, in EMB agar containing a sugar X, a bacterium which has phenotype X$^+$ makes a purple colony and one which has phenotype X$^-$ makes a pink colony. If a population of Lac$^+$ cells is treated with a mutagen that produces Lac$^-$ mutants and the population is allowed to grow for many generations, a few pink colonies will be found on EMB agar plates among a large number of purple ones. If the mutagenized population is plated on the agar immediately after the population has been treated with the mutagen, some colonies appear which are termed *sectored*—they are purple on one side and pink on the other side. Explain the color distribution of these colonies.

4 *E. coli* polymerase I possesses several enzymatic activities. Two important activities are the polymerizing function and the 3'-5' exonuclease. Mutant polymerases have been found which either increase or decrease mutation rates in an organism containing the mutant enzyme. A mutant that increases the mutation rate is called a *mutator*; a mutant that decreases the mutation rate is called an antimutator. The mutator and antimutator activities are usually a result of changes in the ratio of the two enzymatic activities described above. How do you think the ratios change in a mutator and in an antimutator strain?

5 An enzyme has the property that if amino acid 28, which is glutamic acid, is replaced by asparagine in a mutant, all activity is lost. If in this mutant protein, amino acid 76, which is asparagine, is then replaced by glutamic acid, full activity of the enzyme is restored. What can you say about amino acids 28 and 76 in the normal protein?

6 A bacterial colony containing 10^8 cells is taken from an agar surface and suspended in 10 ml of minimal medium. The bacterial strain that was plated was a prototroph. The colony contains no *leu*$^-$ cells. The culture is grown to a cell density of 10^9 cells/ml. If the rate of production of *leu*$^-$ mutants is 2×10^{-9} per cell per generation, will there be any *leu*$^-$ mutants in the final culture? What property of *leu*$^-$ bacteria have you assumed in drawing your conclusion?

7 In what sense is a spontaneous mutation spontaneous?

8 In microorganisms which mutation rates are more easily measured: ability to inability to synthesize proline (Pro$^+$ to Pro$^-$) or the reverse (Pro$^-$ to Pro$^+$)?

9 Which of the following amino acid substitutions would be likely to yield a mutant phenotype if the change occurred in a fairly critical part of a protein: (1) Pro→His; (2) Arg→Lys; (3) Thr→Ile; (4) Val→Ile; (5) Gly→Ala; (6) His→Tyr; (7) Gly→Phe?

10 Several hundred independent missense mutants have been isolated in the A protein of *E. coli* tryptophan synthetase, a protein having 268 amino acids. Fewer than 30 of the positions were represented with one or more mutant. Why do you think that the number of different positions represented by amino acid changes is so limited?

11 The molecule 2-aminopurine is an analogue of adenine, pairing with thymine. It also pairs on occasion with cytosine. What types of mutations will be induced by 2-aminopurine?

12 Nitrogen mustard reacts efficiently with guanine causing ring cleavage and subsequent hydrolysis of the *N*-glycosidic bond. What base-pair change does this cause?

13 Could a hydroxylamine-induced mutant be induced to revert at the same site by hydroxylamine?

14 Can an acridine-induced mutation be induced to revert by treatment with acridine?

15 Can a mutation induced by nitrous acid be induced to revert at the same site by treatment with nitrous acid?

16 Could a mutation induced by hydroxylamine be induced to revert at the same site by treatment with nitrous acid?

17 Which mutagens could cause reversion of a mutation at a hot spot?

18 Two mutations, *his1* and *his2*, are known to map close together within the same gene. Each mutant reverts at a frequency of 10^{-6} revertants per generation. A double mutant has been constructed which carries both mutations in the *his* gene. What would you expect for the frequency of His$^+$ revertants from this strain?

19 A mutation is isolated in a gene *a*. A revertant is selected and the reverse mutation is mapped in gene *b*. The phenotype of this double mutant is A$^+$B$^+$. Separation of the *a* mutation and the new change in gene *b* by recombination show that the alteration in *b* yields organisms with a B$^-$ phenotype. A few other, but not all, mutations in gene *a* revert in this way. Many mutations in gene *b* are independently isolated and revertants of these are isolated. A few of these revertants prove to be mutations in gene *a*. What do these observations indicate

about the relation between the gene products of the two genes within the cell?

20 Two hundred Leu^- mutants of a bacterial strain are examined separately to determine reversion frequencies. Of these, 90 revert at a frequency of 10^{-5}, 98 at 3×10^{-6}, 6 at 3×10^{-11}, and 6 at 10^{-10}. (a) What type of mutant is probably contained in the class whose reversion frequency is 10^{-11}—a single point mutation, a double point mutation, or a deletion? (b) Can you say anything from these frequencies about whether any of the classes of mutations are chain termination mutations?

21 A mutation is created by growth of a cell in medium containing 2-aminopurine. Later, mutant cells are grown in the presence of 2-aminopurine and a revertant is found. Comment on the location of the reverse mutation with respect to the initial mutation.

22 A mutation is induced by exposure to nitrous acid. Revertants are later selected following growth in the presence of 2-aminopurine. Genetic mapping experiments suggest that the reversion event is at the same site as the original mutation (or at least very near). If it is at the identical site, what base pair was present at the site in the wild-type cell?

23 Revertants of temperature-insensitive mutations often prove to be temperature-sensitive. That is, they exhibit a wild-type phenotype at low temperature and mutant at a higher temperature. Explain this phenomenon.

ANSWERS TO PROBLEMS

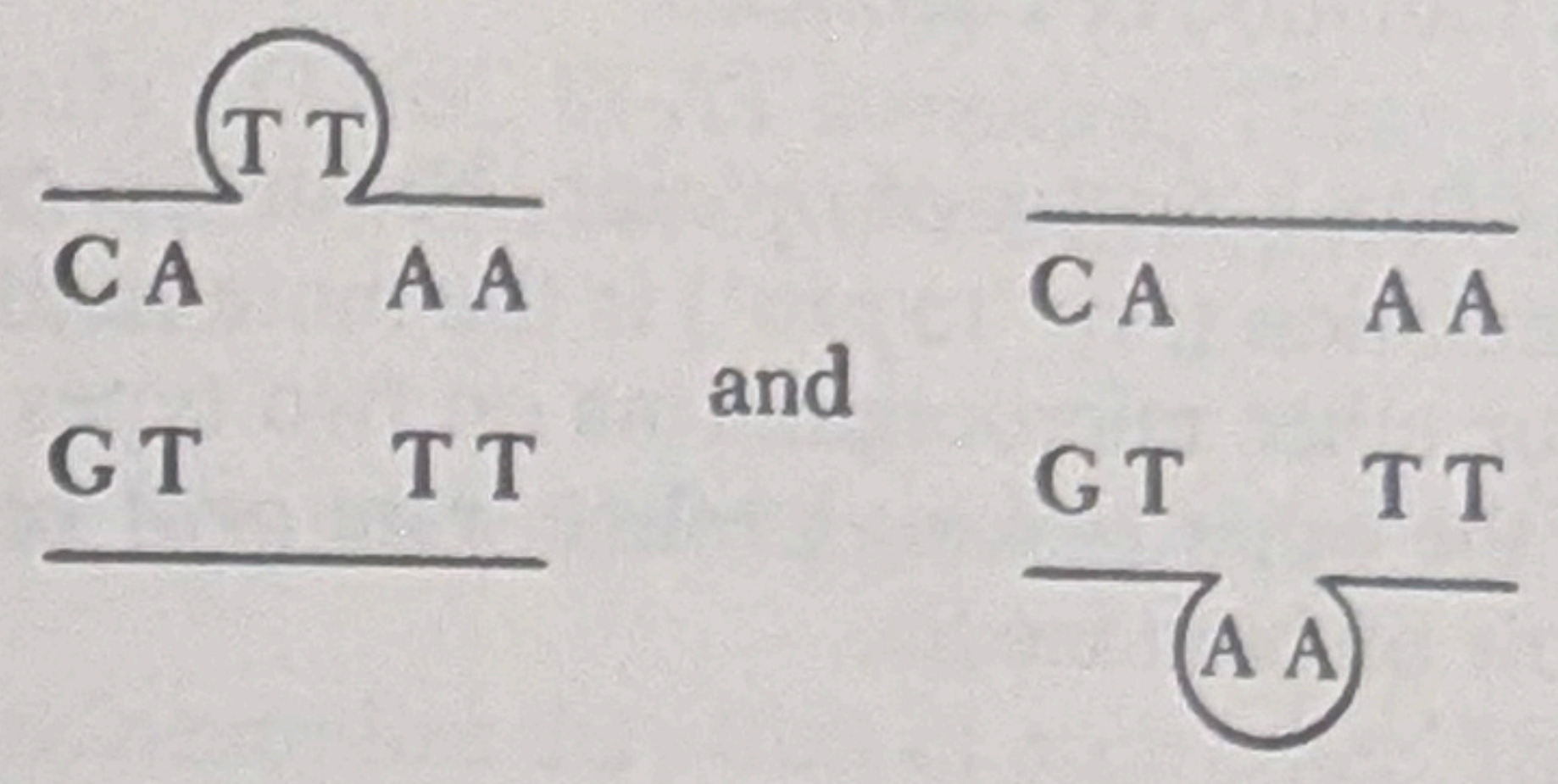
1 (a) First round: One molecule each of

CATTAA	and	CTTTAA
GTAATT		GTAATT

Second round: Three molecules and one molecule, respectively of

CATTAA	and	CTTTAA
GTAATT		GAAATT

(b) First round: One molecule each of



Second round: Two molecules and two molecules, respectively, of

CAAA	and	CATTAA
GTTT		GTAATT

2 This mutant could not be isolated because there is no temperature at which it could grow.

3 The original cell contains an altered base in one strand. Following the first round of DNA replication, which occurs while the bacterium is on the agar surface, two bacteria, one Lac^+ and one Lac^-, result. Since they are adjacent on the agar, the colony has a Lac^+ half (purple) and a Lac^- half (pink).

4. The ratio of polymerizing activity to exonuclease function will decrease in the antimutator since errors will be removed more often.

5 They clearly interact. Since a change in the sign of the charge yields a mutant and a second sign-change in another amino acid yields a revertant, amino acids 28 and 76 are probably held together by an ionic bond.

6 Leu^- cells will certainly be produced but they cannot grow on the medium, which lacks leucine. Mutants will be found as long as they do not die in the absence of leucine. In *E. coli* and most bacteria, amino acid starvation is not lethal. If leucine-starvation were lethal, mutants would not be found unless the cells were collected shortly after the mutant cell formed. Thymine-requiring mutants would probably not be found in such a minimal medium, since they have a half life of about 10 minutes in the absence of thymine. For the leucine case, the number of mutants would be greater if leucine were in the medium since the mutants could multiply as the culture grew. To maximize the number of mutant cells obtained, leucine would have been added to the medium.

7 It does not arise as a result of an action by the experimenter. It could be induced by environmental radiation or chemicals.

8 Ease is of course often a matter of opinion, but in this case there is little doubt that the reverse mutation (*pro*⁻ to *pro*⁺) is the more easily detected. Simply by placing bacteria or other microorganisms on two types of solid medium, one with proline and the other lacking proline, rare *pro*⁺ cells can be detected by their ability to grow on both media.

9 Proline determines the way the polypeptide backbone folds, so any change from proline (change 1) usually yields a mutant phenotype. Arginine to lysine (change 2) is probably ineffective because both have the same charge and nearly the same size. Substitution of threonine, which is weakly charged, by isoleucine, which is quite hydrophobic (change 3), would very likely cause mutation. Valine to isoleucine (change 4), both of which are nonpolar and have nearly the same size, would probably have little or no effect. Glycine to alanine (change 5), both of which are nonpolar, would have little effect unless there was insufficient space for the slightly larger alanine side chain. Histidine to tyrosine (change 6) probably would cause a phenotypic change because of the differences in charge and lack of flexibility, though in many proteins the change might be ineffective. Glycine to phenylalanine (change 7) would probably have an effect because phenylalanine is so bulky that it is likely to cause a change in shape of the protein. For any of the changes mentioned, it is always possible that the shape change is in a part of the molecule that is fairly tolerant of changes.

10 Amino acid substitutions at many positions in a protein have little or no effect on its activity; only certain substitutions occurring at critical positions have a detectable effect.

11 It will be incorporated into DNA opposite a T. In a later round of replication it will pair mostly with T but occasionally with C, leading to a GC pair, at the site of an AT pair. Therefore, the change is an AT-to-GC transition.

12 The site of the removed guanine serves as a template for an A; thus, the change is GC to TA, a transversion.

13 No. Hydroxylamine only causes transitions from GC to AT. A direct reversal would require a reaction with A or T, and hydroxylamine does not react with these bases. However, hydroxylamine could induce a second-site reversion.

14 Yes, because acridines cause both base additions and base deletions.

15 Yes because, nitrous acid causes both AT-to-GC transitions and GC-to-AT transitions.

16 Yes. The hydroxylamine mutation is the result of a GC-to-AT transition, and nitrous acid can react with A and cause an AT-to-GC transition.

17 A hot spot is usually a GMeC-to-AT transition. These can be reverted by nitrous acid, which reacts with A. However, if the hot-spot mutation were a deletion, it could not be reverted.

18 The product of the frequency for each reversion, or 10^{-12}

19 The two mutations neutralize the effect of each other. The two gene products are probably subunits of a multisubunit protein. The mutant subunits interact but a wild-type subunit cannot interact with a mutant subunit.

20 (a) A double point mutation. (b) No. Frequencies give no information about the type of base change.

21 It must be a second-site revertant since 2-aminopurine stimulates AT-to-GC transitions. If the reversion were at the same site, a change from GC to AT would be necessary and that cannot be induced by 2-aminopurine.

22 Since 2-aminopurine can only induce a change from AT to GC and nitrous acid can induce both AT-to-GC and GC-to-AT changes, the original base pair must have been GC.

23 Although a second-site mutation can restore activity, the new interaction that replaces the original intramolecular interaction destroyed by the mutation may not be as strong as the wild-type interaction and hence more easily disrupted by thermal motion.

MORE CHALLENGING PROBLEMS

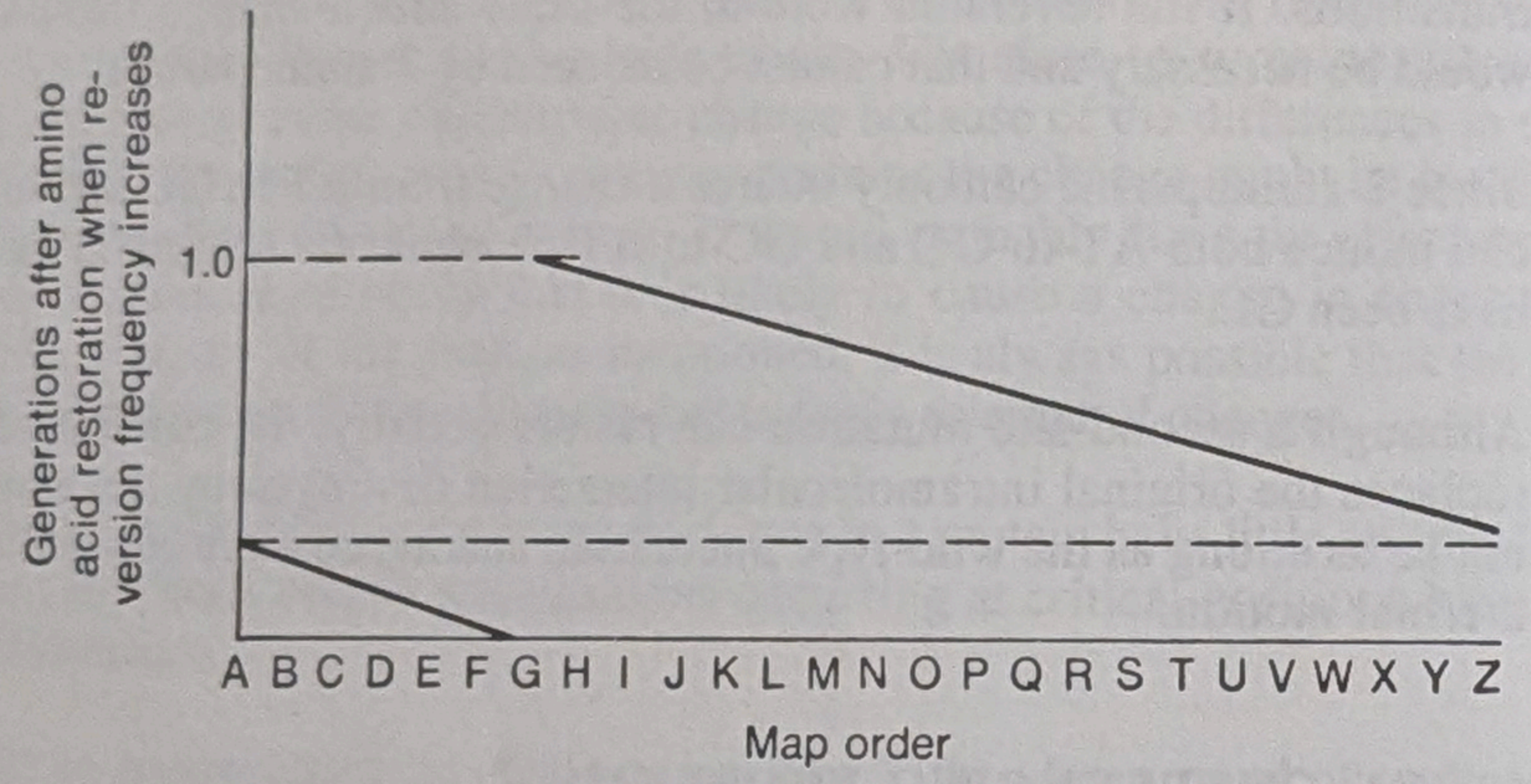
1 A mutation of a bacterial Lac⁺ strain yielding a Lac⁻ colony has been isolated. Several lines of experiments indicate that the mutation resulted from production of a UGA stop codon. Spontaneous revertants are sought and found at a frequency of 10^{-8} per cell per generation and 9 of 10 of them were caused by suppressor tRNA molecules. What do think is the rate of production of suppressor mutations in the original Lac⁺ culture?

2 Since mutation occurs continually in any particular gene, why is it that a population does not become entirely mutant?

3 Assume that a single base is chemically altered in a circular phage DNA molecule contained in a phage particle. If the molecule replicates by the θ mode, half of the daughter molecules will retain the parental base sequence and half will have the altered sequence. If the base causes a mutation, half of the progeny

DNA molecules will be mutant. Consider a phage whose circular DNA is altered in that way and which replicates exclusively by the rolling circle mode. What will be the distribution of mutants among the progeny of individual bacteria infected with a single phage particle?

4 Nitrosoguanidine, a potent mutagen, causes mutations to occur in clusters. That is, a bacterium selected to have a particular mutation often has another nearby mutation. Insight into the mode of action of this mutagen came from studies of bacterial cultures in which DNA replication and cell division were partially synchronized by prolonged amino acid starvation, which prevents initiation but not propagation of a round of replication, followed by amino acid restoration—all cells started their replication cycle at nearly the same time. When such a culture was treated with the mutagen in 1-min pulses, it was found that for each time mutations (actually, revertants) were detected in only a single region of the map. A plot of the time after amino acid restoration *vs* map position is shown below. What does this suggest about the mechanism of mutagenesis by this compound?



5 Consider a synchronously growing and dividing culture of *E. coli*, as described in Challenging Problem 4. In the middle of the replication cycle the cells are given a dose of ultraviolet light. Will the mutations induced by the ultraviolet light be distributed uniformly throughout the map? Explain.

6 S1 nuclease makes a double-strand break at the site of two unpaired bases (a base-pair mismatch). It has been observed that if DNA molecules isolated from a strain of λ phage maintained in a laboratory in the United States are renatured with DNA molecules of a phage that is presumably genetically identical to the first strain but obtained from a laboratory in France, approximately half of the renatured DNA molecules are broken by S1 nuclease. Explain this observation.

7 Most amino acid-requiring bacterial mutants require a concentration of the amino acid of at least 0.020 mg/ml in the agar to achieve a normal colony size (3 mm diameter). A strain of *E. coli* is observed to produce colonies only 1 mm in diameter on a minimal agar yet colonies of normal size if the agar is supplemented with leucine at a concentration of 0.008 mg/ml. Explain.

8 A phage T4 mutant in the *rII* gene is isolated. A revertant of this mutant is isolated that has the wild-type (r^+) phenotype. If this revertant is crossed with authentic r^+ phage and the phage progeny are plated, many plaques are found with the mutant phenotype. Phage stocks are prepared from a large number of these mutant plaques and the stocks are systematically crossed against one another. The plaques are found to fall into two classes A and B defined as follows. If a phage in class A is crossed with a phage in class B, recombinants with an r^+ phenotype are produced. However in a cross between two phage from the same class, no recombinants with an r^+ phenotype are found. Furthermore, it is found that if the phage in either class A or B are crossed with the original mutant, r^+ recombinants are found only in a cross with a class-B phage. Suppose it were known that the original mutation was a deletion of two bases; what kind of mutation or mutations would be present in the class-B phage?

9 For many genes there is no convenient means to enrich a culture for mutants so the mutant fraction is annoyingly small. The following technique has proved to be valuable to isolate such mutants. It makes use of the property that nitrosoguanidine, which is an exceedingly active mutagen, acts primarily in the growing fork. If a mutation in gene *b* is needed, a culture of a strain that has a mutation in a nearby gene (say, *a*) is treated with nitrosoguanidine and revertants of the *a* mutant are selected. Many of these revertants are also *b* mutants. What is the principle underlying this technique?

10 If you wanted to determine whether a particular bacterial gene had mutational hot spots, it is necessary to isolate many mutants and map the mutations. The mutants must be isolated with caution because if a nonmutant culture is allowed to generate mutants by growth through many generations, a mutant produced early in the development of the culture will multiply and yield more mutant cells than a mutant produced late in the growth of the culture. Thus, some mutations will be counted several times and hence distort the data. How might this problem be avoided?

ANSWERS TO MORE CHALLENGING PROBLEMS

1 The mutation frequency in the Lac$^-$ culture is 9×10^{-9}. The probability of production of a suppressor is independent of the presence of a suppressor-sensitive mutation (which the Lac$^-$ mutation is), so the mutation rate in the Lac$^+$ cell is also 9×10^{-9}.

2 If the mutation confers a distinct advantage on the organism, it will become prevalent. If it is disadvantageous, it will be lost rapidly. If it is neutral, it will

continue to increase until the forward mutation rate equals the reverse mutation rate. Few mutations are totally neutral, though.

3 In rolling circle replication one strand is a master strand (the one usually drawn as the "inner" circle). If the base change occurs in that strand,all progeny will be mutant. However, if it occurs in the nonmaster strand (the "outer" strand), only one progeny DNA molecule will be mutant.

4 The time after amino acid restoration at which a gene reverts depends on map position and increases continually from gene *G*. This suggests that the mutagen acts in or near the replication fork and that the replication origin is near *G*.

5 Remember that the major cause of mutagenesis by ultraviolet light is the error-prone synthesis of SOS repair. Since this system functions by allowing replication without editing, all mutations must appear in DNA that replicated after the irradiation. The extent to which the mutations will be uniformly distributed depends on the time it takes for the SOS system to be activated and then deactivated. If activation is very rapid and if it does not last for more than one generation time, all mutations will be in the region of the map corresponding to DNA that had not yet replicated during synchronous growth. (This would, by the way, consist of two regions, since *E. coli* DNA is replicated bidirectionally.) If the system is active for more than one generation, mutations will arise throughout the genome. This experiment, to my knowledge, has not been done.

6 In the course of many rounds of replication in the United States and in France, silent mutations will accumulate. These mutations will not be the same, so that a U.S.-France heteroduplex will have mismatched pairs of bases and thus short single-stranded regions.

7 The strain has a leaky mutation in the *leu* gene; the strain grows but the mutation produces slow growth in the absence of exogenous leucine.

8 Addition of two bases or deletion of one base.

9 If nitrosoguanidine is highly active and acts in only a small region of the DNA, it probably produces mutations in clusters (in fact, it does). Since *a* and *b* are nearby, with high probability a nitrosoguanidine-revertant of an a^- mutation has other nearby mutations, many of which will be in gene *b*.

10 If the mutant is an auxotroph that can survive several generations of starvation for the nutrient, the cells in which the mutations are arising can be grown in a minimal medium lacking the nutrient. An alternative (and the most general technique) is to use several hundred cultures whose size is such that mutations will occur primarily in the last or next-to-last doubling, and start each culture with an aliquot of a single resuspended colony. Each culture will usually contain either no mutants, a single mutant, or the progeny of a single mutant.

CHAPTER 12

TRANSCRIPTION

CHAPTER SUMMARY

Gene expression begins by the enzymatic synthesis of an RNA molecule that is a copy of one strand of the DNA segment corresponding to the gene. This process is called transcription, and the enzyme responsible for it is RNA polymerase. The substrates of RNA polymerase are ribonucleoside triphosphates, and the chemical reaction catalyzed by the enzyme is the same as that used in DNA synthesis. A difference between DNA and RNA polymerases is that RNA polymerases do not need a primer to initiate synthesis. Transcription is initiated when RNA polymerase binds to the polymerase binding site and the polymerization start site of a promoter. In a prokaryote the promoter has several other subregions with which RNA polymerase binds, for example, the sequence TATAAT (the Pribnow box) and TTGACA (the –35 sequence). The promoter-recognition subunit of bacterial RNA polymerase is σ, but whether σ binds to either the Pribnow box or the –35 sequence is not known. In addition, there are other regions, always less than 100 nucleotide pairs upstream from the start site, to which sequence-specific proteins bind; these proteins either strengthen the binding of RNA polymerase to the promoter or facilitate the creation of the conformation of both DNA and the polymerase needed to initiate transcription. Two well-studied factors of this type are the cAMP receptor protein (CRP), which is active only when cAMP is bound to it, and the cII protein. In eukaryotes protein factors are essential for initiation; these bind to the promoter at specific sites and enable RNA polymerase to reach the initiation site. It is unlikely that RNA polymerase can bind to chromatin without the help of these transcription factors. In *E. coli*, RNA polymerase stutters over the first eight (roughly) nucleotides, moving forward and backward, polymerizing and degrading, until a conformational change occurs in the enzyme and the σ subunit is released. Once this occurs, polymerase is locked into the forward motion. However, the forward motion is not constant and pausing frequently occurs, for unknown reasons. Polymerization continues until a termination sequence is reached. There are two types of termination sequences, those requiring the presence of the protein Rho and those that do not. The Rho-independent sequences are characterized by a palindrome and a tract of adenines in the coding strand; thus, at the 3'-OH end the transcript has a stem-and-loop region followed by a tract of U's. No special sequence or DNA structure seem to be present in those regions requiring Rho for termination. In eukaryotes, termination is not understood.

In eukaryotes a variety of DNA sequences upstream from the transcription start site influence the probability of initiation. The most prominent are called upstream sites. They differ from binding sites for ancillary proteins in prokaryotes in that they are usually several hundred nucleotide pairs from the start sites. Furthermore, their precise position is unimportant in that they can be moved by recombinant DNA techniques without seriously affecting initiation. In some cases they can be as far as 1000 nucleotide pairs upstream and still activate transcription. Upstream sites are believed to be binding sites for transcription factors, but other explanations have also been suggested. Regions of chromatin that are excessively sensitive to cleavage by certain nucleases are also present in the upstream region. These are called hypersensitive sites; they may be identical to or near the upstream sites. A most unusual upstream site is the enhancer, a sequence present upstream from some genes that increases the rate of transcription nearly 100-fold. An unexplained feature of the enhancer is that it may not only be moved to various distances from the promoter but can be inverted or even moved downstream from the promoter, without loss of enhancing activity

A variety of antibiotics interfere with transcription. The best-studied is rifampicin, which binds to prokaryotic RNA polymerases and interferes with initiation. α-Amanitin is a useful inhibitor of eukaryotic polymerase II. Actinomycin D inhibits all stages of transcription by binding to DNA.

The product of transcription is an RNA molecule called a primary transcript. In both prokaryotes and eukaryotes, if the primary transcript includes sequences for ribosomal RNA or transfer RNA, the transcript must be cleaved considerably to produce the finished RNA molecule. In prokaryotes the rRNA and tRNA sequences are continuous and must only be cleaved from the transcript. The unusual bases in tRNA molecules are added during or after cleavage by modification of preexisting bases. In eukaryotes there may be intervening sequences that must also be removed; thus, in addition to cleavage, splicing is required. In prokaryotes, when the primary transcript is to serve as messenger RNA (mRNA), it is used directly without modification. In eukaryotes the primary transcript is almost always processed: noncoding sequences called introns are removed, a cap is added directly to the 5' terminus of the primary transcript, and a 3' poly(A) sequence is added. Polymerization often proceeds past the polyadenylation site and a subsequent cleavage event removes an RNA segment from the 3' terminus prior to polyadenylation. Eukaryotes also contain small RNA molecules that exist in the form of ribonucleoprotein particles; these are found both in the nucleus (snRNP) and the cytoplasm (scRNP). Some of the snRNP's participate in processing of primary transcripts, for example, U1 snRNP in splicing and U7 in polyadenylation. The function of the cytoplasmic particles is unknown.

Production of meaningful proteins from a transcript requires RNA splicing because of the presence of the introns, which are noncoding. The coding segments, the exons, must be joined to produce mRNA. A typical intron is terminated at its 5' end by GU and at its 3' end by AG. The 5' site is somehow recognized by the U1 snRNP and cleaved. A series of reactions occur in which the intron 5' terminus is covalently linked to an A within the intron and in which the 3' end of the intron is cleaved. The result is excision of the intron as a lariat structure (with the 3' splice site at the end of the lariat) and linkage of the ends of the adjacent exons.

Prokaryotic mRNA has a half life of only a few minutes, but rRNA and tRNA are long-lived, rarely being subjected to degradation. In eukaryotes a typical lifetime of mRNA is several hours and some molecules survive for days and weeks.

BOLD TERMS

CAAT box, cap, cistron, closed-promoter complex, coding strand, codon, consensus sequence, core enzyme, CRP, cyclic AMP receptor protein, dimethyl sulfate protection method, DNase protection method, downstream, elongation site, exons, enhancer, 40-kd protein, footprinting, 5' splice site, GU–AG rule, heterogeneous nuclear RNA, high-level, Hogness box, holoenzyme, hypersensitive sites, initiation site, intervening sequences, introns, lariat, leader, open-promoter complex, pause, polar mutations, polycistronic mRNA, posttranscriptional modification, pre-tRNA, Pribnow box, primary transcript, processing, processivity, promoter, promoter mutations, reading frame, ribosome, ribozyme, rifampicin, RNA polymerase, rRNA, small cytoplasmic RNA (scRNA), small nuclear RNA (snRNA), scRNP, scyrps, snRNP, snurps, spacers, splicing, strong promoter, TATA, termination factors, TFIIIA, –35 sequence, 3' splice site, transcription, transcription factors, transcription unit, tRNA, tRNA nucleotidyl transferase, U1, upstream, upstream activation sites, upstream sites, weak promoters

ADDITIONAL INFORMATION

Introns are a common feature of eukaryotic primary transcripts. Recently, an RNA molecule synthesized by phage T4 has been shown to contain an intron. This is the first example of a prokaryotic intron; further information is not available at present.

Site selection in splicing reactions is a central question. Clearly, base sequences are important, but their role has not been elucidated in general. Recently, information has become available about site selection in the self-splicing RNA of *Tetrahymena* rRNA (Waring et al. 1986. *Nature,* 321, 133.) In this system an RNA double helix forms by base-pairing between bases surrounding the 5'-splice site and certain bases in the intron (the intron guide sequence). Cleavage occurs at a particular base in this helical region. Proof comes from mutagenesis experiments in which the base sequence has been altered such that the helix forms in a different but nearby region of the RNA. Cleavage occurs in the mutant helix, which causes an incorrect exon-exon joint.

DRILL QUESTIONS

1 What can be said about the relative molar amounts of each base in RNA?

2 What uracil-containing molecule is a precursor for RNA synthesis?

3 What are the chemical termini of a primary transcript?

4 What groups are at the termini of processed RNA in prokaryotes?

5 What groups are at the termini of processed rRNA and tRNA in eukaryotes?

6 What is the relation between a primary transcript and an mRNA molecule in prokaryotes?

7 What sequences or groups are at the termini of mRNA in eukaryotes?

8 What common sequences are found in the 3'-terminal segment of prokaryotic mRNA?

9 In what way, if at all, does the chemistry of polymerization by RNA polymerases differ from that by DNA polymerases?

10 What is the direction of synthesis of RNA?

11 If a DNA-RNA hybrid were made, would the strands be parallel or antiparallel?

12 How must the statement, "both DNA strands cannot serve as a template for transcription," be modified to make it true?

13 Define coding strand and antisense strand.

14 How many subunits are in *E. coli* RNA polymerase holoenzyme and which one is responsible for promoter recognition?

15 Phage-specific RNA polymerases are always small enzymes consisting of a single subunit. Why are the bacterial polymerases larger and more complicated?

16 In eukaryotes, which polymerase synthesizes transcripts that will ultimately become mRNA?

17 What information is gained by footprinting?

18 With respect to a promoter, in what direction along a DNA strand is upstream? Is an upstream region transcribed?

19 What two regions are common to most prokaryotic promoters?

20 What is the consensus sequence for the Pribnow box?

21 What consensus sequence is present in a large number of eukaryotic promoters? What sequence is common but less frequent?

22 RNA polymerase can bind to DNA with or without causing separation of the DNA strands. What terms describe the DNA-enzyme complex in the two states?

23 How does the binding of an RNA polymerase to a promoter in prokaryotes differ from that in eukaryotes?

24 In what part of a eukaryotic promoter does RNA polymerase II actually bind to the DNA?

25 What stage of RNA polymerization is inhibited by rifampicin in prokaryotes?

26 At which ends of eukaryotic mRNA are the cap and the poly(A) tail?

27 During chain elongation roughly how many RNA bases are paired to DNA?

28 During chain elongation how is *E. coli* RNA polymerase related to the holoenzyme?

29 In *E. coli* how is the core enzyme of the holoenzyme related to the core enzyme when chain elongation is occurring?

30 Does an RNA polymerase core enzyme normally bind again to the σ subunit that it has released?

31 Does *E. coli* RNA polymerase, like DNA polymerases I and III, move continuously along the DNA?

32 Do pause sites serve any known function?

33 What types of primary transcripts are processed in prokaryotes?

34 In prokaryotes and in eukaryotes are rRNA molecules contained in primary

transcripts containing only rRNA, and tRNA in primary transcripts containing only tRNA?

35 Both rRNA and tRNA precursors are spliced in forming the finished RNA molecules. However, only in forming mRNA is the excised RNA called an intron. What feature defines an intron?

36 What base sequences are typically found at the 5' and 3' ends of introns?

37 Which terminus of an intron is cleaved first?

38 Are any eukaryotic primary transcripts known that lack introns?

39 Roughly, what are the relative sizes of introns and exons?

40 RNase P is one of the enzymes involved in production of tRNA from the primary transcript. How does it differ in its mode of action from other processing enzymes?

41 What sequence is present at the 3' terminus of all completed tRNA molecules?

42 What is a typical lifetime of a prokaryotic mRNA molecule?

43 In what way(s) do the templates for transcription in prokaryotes and eukaryotes differ?

44 What is meant by heterogeneous nuclear RNA?

45 How do the primary transcripts of prokaryotic and eukaryotic tRNA molecules differ?

46 How do eukaryotic polymerase-II and polymerase-III promoters differ?

47 How do upstream sites in eukaryotes differ from binding sites for particular proteins in prokaryotic promoters?

48 What are the possible functions of upstream sites in initiation of transcription in eukaryotes?

49 What is meant by a hypersensitive site?

50 Apart from their ability to stimulate transcription, by what two features do enhancers differ from all other elements that affect transcription rate?

51 In higher eukaryotes roughly what fraction of each primary transcript is discarded in the formation of mRNA?

52 RNA is precipitated by acid. Is acid-precipitable material labeled by growth of cells in a medium containing [^{3}H]uridine probably exclusively RNA?

ANSWERS TO DRILL QUESTIONS

1 Nothing. They can have any value.

2 Uridine triphosphate (UTP).

3 A 5'-triphosphate and a 3'-OH.

4 A 5'-monophosphate and a 3'-OH group.

5 A 5'-monophosphate and a 3'-OH group.

6 They are the same.

7 A 5' cap and a 3' poly(A) sequence.

8 A stem-and-loop and a tract of U's.

9 Not at all.

10 From the 5' end to the 3' end.

11 Antiparallel.

12 In a given DNA segment both DNA strands cannot serve as a template for transcription. Actually there are a few known exceptions to this rule in which transcribed regions overlap slightly.

13 A coding strand is a segment of a DNA strand that is copied by RNA polymerase. An antisense strand is a DNA strand that is complementary to a coding strand.

14 Five; σ subunit.

15 The phage polymerase need only recognize a small number of different promoters, whereas the bacterial enzyme must recognize hundreds.

16 RNA polymerase II.

17 The technique indicates the DNA base sequence to which a protein is bound.

18 In the 3' direction. No, it comes before the transcribed segment.

19 The Pribnow box and the –35 sequence.

20 TATAAT.

21 TATA is fairly common; CAAT is less common.

22 Separation, open-promoter complex; no separation, closed-promoter complex.

23 In prokaryotes RNA polymerase can usually bind directly to a promoter without the help of accessory proteins. In eukaryotes RNA polymerases are directed to promoters by promoter-specific proteins.

24 Probably only at or very near the transcription start site.

25 Chain initiation.

26 The cap is at the 5' end; the poly(A) tail is at the 3' end.

27 About 12.

28 It lacks the σ subunit.

29 The core enzyme probably undergoes a conformational change when the σ subunit falls off.

30 No. σ subunits are acquired randomly from an internal pool of free subunits.

31 No. It slows down and perhaps stops in regions called pause sites.

32 It is not clear, but some are probably needed for Rho-independent termination.

33 Only those containing tRNA and rRNA molecules.

34 In prokaryotes primary transcripts often contain both rRNA and tRNA sequences. In eukaryotes 5S rRNA is in a transcript containing no other molecules, and the other three rRNA molecules are in a separate transcript; no tRNA sequences are present in these transcripts.

35 An intron interrupts a coding sequence.

36 5'-AG and GU-3'.

37 The 5' terminus.

38 Yes, a few in higher eukaryotes and many in yeast.

39 There is no rule; introns can be both larger and smaller than exons.

40 It is a ribonucleoprotein and the enzymatic activity resides in the RNA. That is, it is a ribozyme.

41 CCA-3'-OH.

42 A few minutes.

43 In prokaryotes the template is free DNA; in eukaryotes, it is chromatin.

44 The collection of primary transcripts and partially processed RNA in the nucleus of a eukaryote.

45 In eukaryotes the pre-tRNA frequently contains an intervening sequence in the tRNA sequence; the prokaryotic tRNA sequence is uninterrupted.

46 Polymerase-III promoters are downstream from the transcription start site, whereas those of polymerase II are upstream.

47 (1) In eukaryotes the upstream sites are further from the transcription site. (2) Upstream sites can be moved somewhat upstream or downstream without eliminating activity of the promoter; binding sites in prokaryotic promoters are inactivated by even a small relocation.

48 They may "loosen" the chromatin structure and/or serve as binding sites for transcription factors.

49 A region of eukaryotic chromatin that is much more easily cleaved by certain DNases than the bulk of the DNA.

50 It is active both upstream and downstream from a transcription start site and it functions in both orientations.

51 Usually more than half and up to 90%.

52 No, because labeled U can be converted to T and incorporated into DNA, which is also acid-precipitable. Treatment with a DNase prior to precipitation solves the problem.

PROBLEMS

1 (a) Write down the two RNA sequences which could conceivably result from complete transcription of a DNA molecule that has the following sequence in one of the strands—5'-AGCTGCAATG-3'. Indicate the 5' and 3' ends of each

transcript. (b) In an *in vitro* system would both transcripts be made?

2 RNA polymerase lacks the nuclease activities of DNA polymerase. Name two activities possessed by RNA polymerase and not by DNA polymerases.

3 How does the DNase protection method work and what information does it yield?

4 What information is provided by the dimethyl sulfate protection method?

5 A chromatographic column in which oligo-dT is linked to an inert substance is useful in separating eukaryotic mRNA from other eukaryotic RNA molecules. On what principle does this column operate?

6 What is probably the main reason that transcription units are more complicated in eukaryotes than in prokaryotes?

7 What parts of a mRNA molecule are not translated?

8 What is a transcription unit and is it the same thing as a gene?

9 Describe two types of promoter mutations.

10 In what ways are the templates for transcription in prokaryotes and eukaryotes different?

11 What RNA polymerases exist in eukaryotic cells? What are their locations and their functions *in vivo*?

12 Without carrying out any sequencing analyses what indication might you have that all promoters do not have the same overall base sequence?

13 When the RNA polymerase reaction is carried out *in vitro*, it is found that at physiological ionic strength fewer regions of a particular DNA molecule are transcribed than when the ionic strength is very low. Furthermore, at low ionic strength, the regions described differ markedly from those observed to be transcribed in vivo. Explain.

14 Radioactive RNA can be prepared by enzymatic synthesis using either α-^{32}P-labeled nucleoside triphosphates (NP*PP, in which the * indicates that the P to its left is radioactive) or γ-^{32}P-labeled molecules (NPPP*). If the radioactivity of the α- and γ-labeled molecules is the same, the total radioactivity incorporated as acid-insoluble material (that is, RNA, but not individual nucleotides) will be greater with the α-labeled precursors. Explain and state what aspect of RNA synthesis is actually measured with the different types of precursors.

15 The rate of initiation of new RNA strands *in vitro* is greater with a superhelical template than with a linear DNA molecule or with a nicked circle. Propose an explanation for this phenomenon.

16 Some DNA molecules have complementary single-stranded ends which allow circularization by means of base-pairing of these segments. DNA ligase can be used to seal the single-strand breaks, yielding a covalent circle without superhelical turns. If RNA polymerase is bound to the DNA before the joining of the end segments and ligation and then the RNA polymerase is removed, the DNA circle has superhelical turns.What does this say about the binding of RNA polymerase to DNA?

17 How does one determine the first base transcribed in a promoter sequence?

18 The Pribnow box sequence is an example of what is termed a "consensus sequence"—namely, a base sequence from which other sequences having similar functions can be obtained by changing only one or two bases. (a) What is the evolutionary significance of such a sequence? (b) What is the biochemical significance of a conserved base, such as the conserved T in the Pribnow box? (c) What biochemical differences might you expect between sequences that differ only slightly in the nonconserved bases?

19 A particular sequence containing 6 base pairs is located in 10 different organisms. The observed sequences are ACGCAC, ATACAC, GTGCAC, ACGCAC, ATACAC, ATGTAT, ATGCGC, ACGCAT, GTGCAT, are ATGCGC. What is the consensus sequence?

20 Modern genetic engineering techniques allow the movement of genes from one organism to another. Placing eukaryotic genes into *E. coli* has been of special interest. It has been observed that transferred animal genes are not always transcribed in *E. coli* yet transcription occurs much more frequently in yeast. Why?

21 Why is the SSB protein not needed by RNA polymerase?

22 If purified *E. coli* RNA polymerase holoenzyme is added to a particular DNA molecule in the presence of the four necessary nucleoside triphosphates, RNA is made. Analysis of the size of the RNA chains made shows a collection of mRNA molecules having a wide range of sizes, and the size distribution is nearly continuous. If a small amount of a cell extract (demonstrated to be free of RNase activity) is added, smaller fragments of discrete size are observed. What is probably present in the extract?

23 A reaction mixture contains a unique DNA segment (containing only one complete gene), *E. coli* RNA polymerase, and four γ-^{32}P-labeled nucleoside triphosphates (in order to have one ^{32}P per RNA molecule, see Problem 14). RNA is isolated shortly after the reaction begins, and the lengths of all RNA molecules

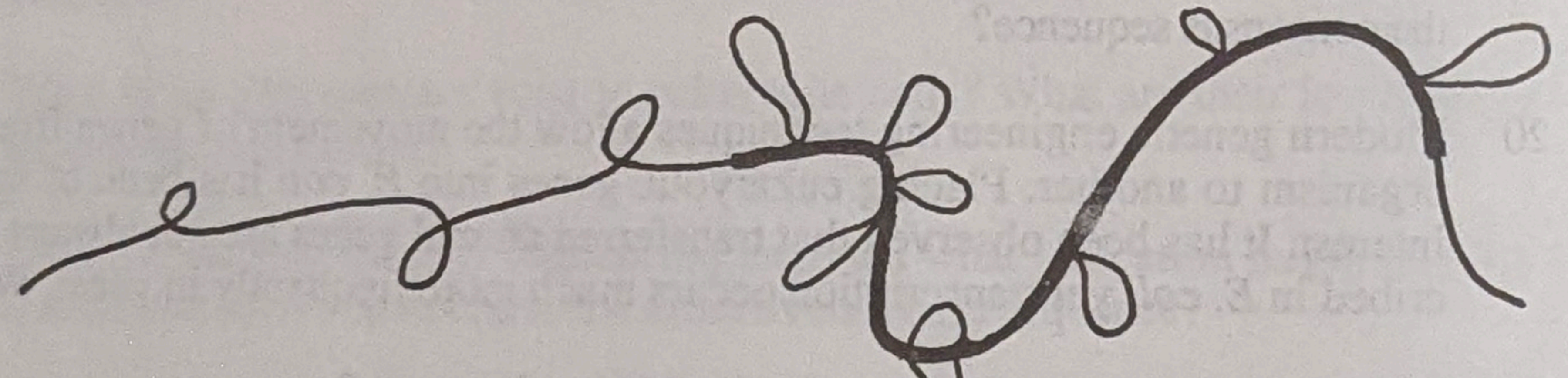
in the mixture are determined by gel electrophoresis, using autoradiography to visualize the bands. The longest molecule observed contains 92 nucleotides. Bands are seen in all positions, though in general there is less radioactivity in the bands corresponding to large molecules than to small molecules. (a) Why? (b) In addition, there is a large stepwise decrease in the amount of radioactivity in each band containing more than 28 nucleotides than in each band containing fewer than 28 nucleotides, and another large stepwise decrease at the band corresponding to 55 nucleotides. Explain these stepwise decreases.

24 In prokaryotes how is the 1:1:1 ratio of 5S, 16S, and 23S rRNA maintained?

25 An RNA molecule is isolated having a 3'-OH terminus and a 5'-P terminus. What information does this fact provide?

26 What are the three steps in the production of a tRNA molecule from a primary transcript?

27 A plasmid is prepared by genetic engineering to contain a gene that is a copy of the mRNA encoding the product of a particular gene. The plasmid DNA is isolated, denatured, hybridized with the primary transcript of the gene. The electron micrograph drawn schematically in the figure below is obtained. (Thin and thick lines refer to single- and double-stranded nucleic acid respectively.) How many introns are there in the gene?



28 Consider the processing of *Tetrahymena* pre-rRNA. (a) In what sense is the reaction not enzymatic? (b) In what sense is it not even catalyzed?

29 By genetic engineering genes can be moved from one organism to another. By coupling a eukaryotic gene to an *E. coli* promoter transcription of the gene will occur but the proper protein is usually not synthesized. However, synthesis of the correct protein frequently occurs if the gene is transferred to yeast and coupled to a yeast promoter. Why?

ANSWERS TO PROBLEMS

1 (a) PPP-5'-AGCUGCAAUG-3' and PPP-5'-CAUUGCAGCU-3'. (b) If the system required definite promoter sequences, probably no RNA would be made. If conditions were chosen for nonspecific initiation, both RNA molecules would be made.

2 It is self-priming (that is, it does not need a primer), and it can unwind DNA without the need of a helicase.

3 DNA that is bound to protein is resistant to pancreatic DNase. Thus, if a complex of DNA and RNA polymerase is exposed to DNase, only the segment of DNA over which the bound RNA polymerase extends will survive enzymatic digestion. The number of nucleotides and the base sequence in this region can be determined by simple additional measurements.

4 Bases in close contact with a protein can be identified, because such bases are resistant to methylation.

5 Eukaryotic mRNA molecules contain a terminal poly(A), which can hydrogen-bond to the dT in the column under renaturing conditions. Thus, the mRNA will be retained by the column and other RNA species will pass through. The mRNA can be eluted from the column simply by heating.

6 Eukaryotic DNA is tightly bound to histones, so the bases are not readily available to RNA polymerases.

7 Leaders, spacers, and the unnamed region following the last stop codon of an mRNA are untranslated regions.

8 A transcription unit is a section of DNA extending from a promoter to an RNA polymerase termination site. It is usually not a gene, but typically includes many genes.

9 Promoter-up mutation: a new promoter is created or the activity of an existing promoter is enhanced; promoter-down mutation; a natural promoter is inactivated.

10 In eukaryotes the DNA is arranged in nucleosomes.

11 RNA polymerase I makes rRNA in the nucleolus. RNA polymerases II and III are nuclear; the former makes mRNA and the latter makes tRNA and 5S RNA.

12 The rates of initiation of synthesis of all mRNA molecules in a particular organism differ from one gene to the next.

13 At low ionic strength RNA polymerase binds at random to numerous charged sites rather than to specific promoter sequences. At high ionic strength the charges are shielded and binding occurs only to promoters.

14 In polymerization a pyrophosphate is removed, leaving only the α phosphates in the RNA. The γ label can persist only at the 5' terminus. Thus, incorporation of γ-labeled radioactivity is a measure of initiation, and incorporation of α-labeled radioactivity measures overall synthesis.

15 RNA polymerase must come into contact with bases, which means that some unwinding of the DNA is necessary. Recall that supercoiled DNA is partially underwound and that single-stranded regions have a longer lifetime than in nonsupercoiled DNA. Thus, at any given point in the DNA the bases in supercoiled DNA will be accessible to the polymerase for a greater fraction of the time.

16 RNA polymerase must cause either a slight underwinding or overwinding of the segment of a DNA molecule to which it is bound. Whether it causes under- or overwinding would result in either a negative or a positive superhelix.

17 The promoter sequence is first determined. The 5'-terminal sequence of the transcript is then obtained (4 to 5 bases are sufficient) and the complement to this sequence is found in the promoter sequence.

18 (a) It is probably a sequence that existed at very early times and from which others were derived by mutation. Furthermore, it indicates that the biochemical system that uses the sequence existed very long ago. (b) It is essential for some stage of promotion, in particular, binding of RNA polymerase or initiating polymerization. (c) The rates of initiation may differ slightly. However, it is difficult to be certain of this point, because the rate of initiation is mainly determined by the –35 region.

19 The number of times each of the four bases appears at each position is noted. The maximum value (as a percentage) is indicated as a subscript in the consensus sequence that follows: $A_{80}T_{70}G_{80}C_{90}A_{80}C_{70}$.

20 *E. coli* is a prokaryote and its RNA polymerase does not recognize all eukaryotic promoters. Yeast, on the other hand, is a eukaryote.

21 Only a few base pairs are broken during transcription. The open region is smaller than the size of RNA polymerase, which is apparently able to break a few base pairs by itself, as it moves along the helix.

22 The Rho factor.

23 (a) Since each molecule contains only a radioactive terminus, there will always be less radioactivity in the bands representing the longer RNA molecules because many molecules are in the act of synthesis at the time the reaction is terminated. (b) The stepwise decreases represent pause sites.

24 One copy of each is present in the primary transcript and each is cleaved out.

25 Since the triphosphate of the primary transcript is absent, the molecule has been processed. The processing could be extensive or as simple as triphosphate hydrolysis but there is no way of knowing from the information given.

26 Formation of (1) the 3'-OH and (2) the 5'-P termini, and (3) modified bases.

27 8.

28 (a) No protein is required. (b) No molecule emerges from the reaction unchanged, which is the strict definition of a catalyst.

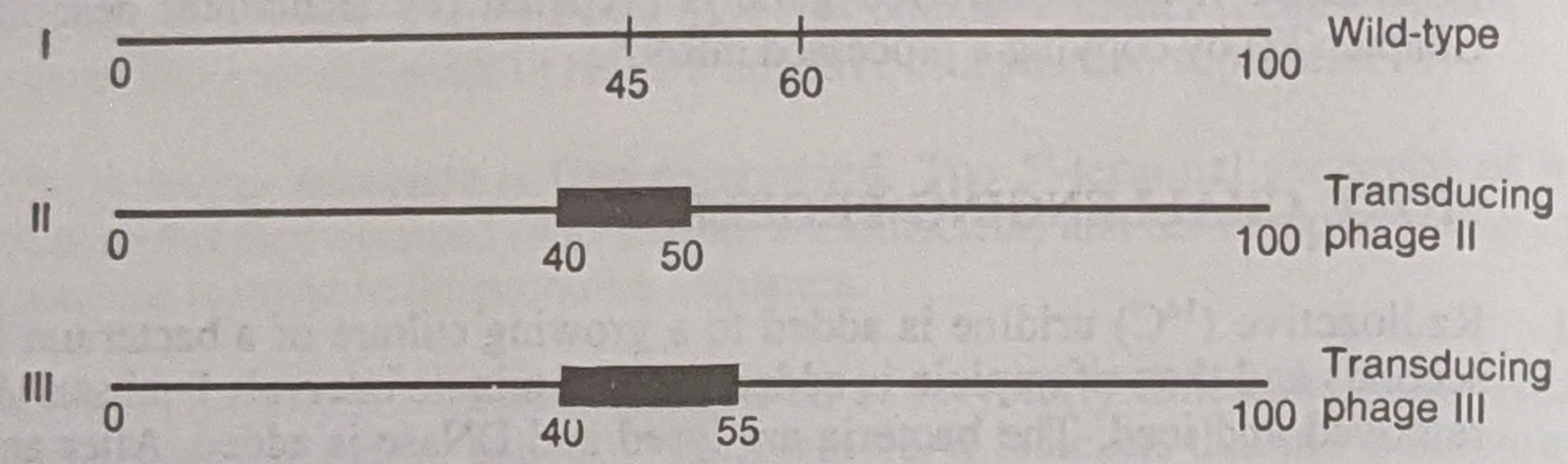
29 *E. coli* cannot remove the introns from the mRNA. Yeast can. This problem can be avoided if the transferred gene is prepared (by techniques described in Chapter 22) by copying a processed mRNA.

MORE CHALLENGING PROBLEMS

1 Radioactive (^{14}C) uridine is added to a growing culture of a bacterium for 10 seconds and then rifampicin is added. At two-minute intervals 1-ml samples are removed and iced. The bacteria are lysed and DNase is added. After an incubation time of one hour to ensure complete DNase digestion, each sample is divided into two portions. One portion is precipitated and the other is hybridized with the DNA of a phage carrying the bacterial *blh* gene. The data are shown in the table below. Plot the data on semilog paper and assume that mRNA decay is first-order. (a) What fraction of the labeled RNA is stable RNA (rRNA and tRNA)? (b) What is the half-life of the total bacterial mRNA? (c) What is the half-life of *blh* mRNA? (d) What was the purpose of the DNase treatment?

Time in minutes	*Acid precipitable radioactivity, cpm*	*Hybridizable radioactivity, cpm*
t = 0	50,240	1820
t = 2	35,050	1375
t = 4	27,510	1060
t = 6	23,725	918
t = 8	21,850	720
t = 10	20,920	561
t = 12	20,460	431
t = 14	20,230	310
t = 16	20,110	248
t = 18	20,010	210

2 Messenger RNA is usually measured by various hybridization techniques, that is, through the formation of a DNA-RNA hybrid by incubation of the mRNA with single-stranded DNA. In one technique, single-stranded DNA is adsorbed to a nitrocellulose filter, radioactive RNA is added, and then the mixture is incubated under conditions leading to hybridization. Free RNA does not bind to the filter, so the amount of radioactivity bound to a washed filter is a measure of the amount of RNA hybridized. In one experiment radioactive mRNA was isolated from a phage-infected cell one minute after infection and then hybridized with three different DNA molecules separately adsorbed to filters. The DNA molecules are shown in the figure below, in which numbers refer to distance from the left end of the DNA, on a scale of 0 to 100.



Wild-type phage (I) were used in the infection. The shaded regions in phage DNA molecules II and III represent bacterial DNA carried by the phage. Following hybridization with each type of DNA on two separate filters, one filter of each type was incubated with a ribonuclease (+RNase) that digests single-stranded RNA but not hybridized RNA; the other filter was untreated (–RNase). The data in the table below were obtained. From which region of the wild-type DNA was the mRNA transcribed?

DNA on filter	*cpm on filter* *–RNase*	 *+RNase*
I	1,250	1,245
II	1,260	820
III	1,242	418

ANSWERS TO MORE CHALLENGING PROBLEMS

1 (a) The graph shows that 20,000 cpm does not decay. This is the stable RNA or 20/50 = 40% of the total. (b) To obtain this number, subtract 20,000 from each value and replot on semi-log paper. The half-life is 2 min. (c) These numbers can be used without correcting for stable RNA, so the half-life is 6 min. (d) Uridine is ultimately converted not only to rUTP but to dUTP, which then forms TTP and is incorporated into DNA. The DNase treatment insures that all acid-insoluble radioactivity is in mRNA.

2 The region from 45 to 60.

CHAPTER 13

Translation I: The Information Problem

CHAPTER SUMMARY

The flow of information from a gene to its product is DNA to RNA to protein. The DNA-to-RNA step, transcription, was described in Chapter 12. The RNA-to-protein stage has two aspects: the system that translates a base sequence into an amino acid sequence, and the mechanism of joining of the amino acids. This chapter is concerned with the translation problem.

The genetic code is the collection of base triplets that correspond to the amino acids. There are 64 codons, of which 61 correspond to amino acids. The remaining three codons are signals for stopping protein synthesis. These codon assignments apply to almost all organisms, and the code is said to be universal. Known exceptions occur in organelles, such as mitochondria and chloroplasts, and in a few microorganisms. However, even these exceptional codes differ from the universal code in only a few codon assignments. Since the 61 codons specifying amino acids match with only 20 amino acids, the code is highly redundant. The first two bases in a codon are amino-acid specific, but very often the identity of the third base does not affect the amino acid assignment; for example, ACX corresponds to threonine, where X can be any base. Only two amino acids are represented by a single codon, namely, methionine (AUG) and tryptophan (UGG).

The methionine codon AUG is the most common start codon. Since an AUG may be within a coding sequence, that is, proteins may contain methionine at internal positions, some mechanism is needed to identify a particular AUG as a start signal. In prokaryotes an initiating AUG is identified by being adjacent to the Shine-Dalgarno sequence, an RNA base sequence that is complementary to a sequence in 16S rRNA. In polycistronic mRNA an AUG can signal start if it follows shortly after a stop codon of the previous protein. In eukaryotes the AUG nearest the 5' terminus of the mRNA is the usual start codon. However, nearby base sequences can influence the efficiency of initiation, and hence the rate of synthesis of a particular protein, and in some cases an adjacent base sequence can prevent initiation; indeed in some organisms a second or third AUG serves as the start codon in some mRNAs.The stop codons are UAA, UAG, and UGA. Stop codons frequently occur in pairs to insure termination.

Translation utilizes two sets of macromolecules: tRNA molecules, which read the codons in RNA, and the aminoacyl synthetases, enzymes that attach particular amino acids to the appropriate tRNA. The tRNA molecules are small RNA molecules, containing 73–93 nucleotides. About half of the bases are hydrogen-bonded, which gives the molecule a cloverleaf shape with three stem-and-loop regions, a fourth stem, and, in some molecules, a hairpin branch (the extra arm). The amino acid is covalently linked to the 3' terminus of the free stem, the acceptor stem. The middle loop, the anticodon loop, is responsible for codon recognition, and contains a nucleotide triplet that can hydrogen-bond with one or more codons. Other loops are involved in ribosome binding. These loops and other features of each tRNA molecule enable a particular aminoacyl synthetase to link an amino acid to a tRNA molecule with the correct anticodon loop. The aminoacyl synthetases occasionally make mistakes but possess a steric mechanism (the double-sieve) for removing an amino acid that has been incorrectly linked to tRNA.

The codons in mRNA are read sequentially. Reading consists of base-pairing of the anticodon to a codon. This pairing occurs when the mRNA is bound to a ribosome. Two tRNA molecules, each bearing a covalently linked amino acid, are place side-by-side on the ribosome, resulting in alignment of two amino acids in the same order as the corresponding codons in the mRNA. Bases other than A, C, G, and U are found in the anticodon. Inosine (I, hypoxanthine riboside) is a prominent base in the first position of some anticodons. Since the base-paired region is only three base pairs in length, the constraints that demand that only AT and GC pairs exist in DNA are not applicable to codon-anticodon binding. In fact, the first anticodon base can form several different base pairs in a codon-anticodon region, and these may be pairs other than AU and GC, for example, IA, IU, IC, and GU. This lack of strict base-pairing specificity, which is called wobble, means that a single tRNA molecule can respond to several codons. The wobble pairings account for the redundancy of the code and explain the number of codons for each amino acid.

Initiation of protein synthesis utilizes a special initiator tRNA molecule, whose structure differs slightly from other tRNAs. In prokaryotes the initiator tRNA is charged with *N*-formylmethionine instead of methionine. The formyl group is added to the methionine after linkage to the tRNA and is removed from the polypeptide chain after protein synthesis is initiated. In eukaryotes formylation does not occur, and the initiating amino acid is methionine.

Mutations can occur in tRNA genes and give rise to altered tRNA molecules. If the mutation is in the anticodon, a tRNA can result in which a particular amino acid is no longer matched with the correct codon. Such altered tRNAs are called suppressor tRNAs. If they can insert an amino acid at a site of a stop codon, they are nonsense suppressors. The presence of a nonsense suppressor can counteract the effect of a chain-termination mutation and allow a prematurely terminated mutant protein fragment to be completed. If the amino acid replacement yields a functional protein, the effect of the chain-termination mutation on the organisms is suppressed. Suppressors exist for each stop codon. The existence of a UAG or UGA suppressor tRNA does not affect normal termination since these codons usually are adjacent to a a second stop codon. A UAA suppressor imposes some problems on the organism containing it since UAA frequently occurs singly as a stop codon signaling normal termination. An efficient UAA suppressor would result in lack of termination of most proteins. Hence, UAA suppressors are weak in that the anticodon binds quite weakly to the UAA codon; a few mutant proteins are not terminated but most are. Missense suppressors cause substitution of one amino acid for another. These are necessarily weak, because a strong missense suppressor would cause alterations in every protein in the organism.

Base sequences are read sequentially, that is, the code is a nonoverlapping code. However, coding sequences occasionally overlap in the small viruses; that is, a single sequence is read in two different reading frames. This can occur by using two distinct start codons for initiation of translation of a particular sequence. If both are in the same reading frame, the two proteins will have a region of common amino acid sequence or one protein will be a fragment of the other. If they are in different reading frames, their amino acid sequences will be completely different. In some cases, a primary transcript is spliced in several ways, so a single start codon can lead into different base sequences. The existence of overlapping genes enables an organism to increase the efficiency of utilization of its DNA. That is, if there are no overlapping genes, the maximum number of amino acids encoded in a particular DNA cannot exceed 1/3 the number of nucleotides. With overlapping genes, the number can be considerably greater.

BOLD TERMS

A site, acceptor stem, amber codon, amino acid attachment site, aminoacyl synthetase, aminoacylated, aminoacyl-tRNA, aminoacyl-tRNA synthetase, anticodon, anticodon deaminase, anticodon loop, base substitution, CCA, chain termination, charged, codon, cognates, degenerate, DHU loop, double-sieve mechanism, extra arm, genetic code, leader, missense, nonoverlapping code, nonsense, ochre codon, opal codon, overlapping code, overlapping genes, P site, photoactivated cross-linking, read-through, recognition region, redundant, ribosomes, RNase-protection, scanning hypothesis, Shine-Dalgarno sequence, spacer, suppress, suppressor mutant, suppressor-sensitive mutations, suppressor tRNA, synonyms, tertiary base pair, TψC loop, transfer RNA, translation, triplet code, tRNA, weak suppressor, wobble hypothesis

ADDITIONAL INFORMATION

It has been pointed out several times that suppression is rarely perfect. An interesting collection of temperature-sensitive suppressors of amber mutations has been isolated. These suppress amber mutations at 30 C but not at 42 C. (Note that these are not suppressors of Ts mutations.) Most of these suppressors are tRNA molecules in which the acceptor stem has been altered, which allows the tRNA to undergo significant changes in shape at 42 C. The Ts suppressors have not been studied in detail but are instead used as an experimental tool. For instance, synthesis of a protein containing an amber mutation proceeds normally at 30 C in a bacterium having a Ts suppressor. However, synthesis can be immediately stopped merely by raising the temperature sufficiently. Of even greater value is the fact that the number of functional molecules depends on temperature, so that the effect of the internal concentration of a molecule can be studied merely by growth of the cells at several temperatures; in this way the internal concentration of a protein can be varied systematically.

DRILL QUESTIONS

1 In terms of its role in translation, what is a ribosome?

2 How many codons could be contained in a four-letter code?

3 What amino acids correspond to the codons UUG and AAU? (Use the table; these do not have to be memorized.)

4 Poly(U) encodes polyphenylalanine. If a G is added to the 3' terminus, the polyphenylalanine will have another amino acid at its C terminus. What is the amino acid, and will the same amino acid be added at the C terminus of polyphenylalanine if the G is added to the 5' terminus?

5 What polymer would be made by the polynucleotide poly(GC)?

6 What two amino acids have only one codon?

7 What are the three stop codons?

8 What is the principal start codon and to what amino acid does it correspond?

9 Which two amino acids can be placed in a polypeptide chain at an AUG codon in prokaryotes. How does the situation differ in eukaryotes?

10 How do prokaryotes and eukaryotes differ in the methods for selecting an AUG site as a start codon?

11 How many tRNA molecules respond to AUG?

12 What are the names of the three loops in all tRNA molecules?

13 What sequence of three bases is at the 3' end of all tRNAs?

14 Is the amino acid linked to the 5' or the 3' terminus of a tRNA?

15 How is an amino acid altered before it is attached to a tRNA?

16 What is meant by the stems, loops, arms, and tertiary base pairs of tRNA?

17 What are the three types of hydrogen bonds in tRNA?

18 Inosine is a nucleoside. What base does it contain?

19 What normal bases can pair with inosine?

20 What base is never found in the first (5') position of an anticodon?

21 If a codon were written PQR, would the anticodon be P'Q'R' or R'Q'P', in which a ' denotes a complementary base?

22 What amino acids are bound to charged $\mathrm{tRNA^{Leu}}$ and to seryl-$\mathrm{tRNA^{Leu}}$?

23 Which of the following properties are essential for the function of a tRNA molecule: (1) recognition of a codon; (2) recognition of an anticodon; (3) ability to distinguish one amino acid from another; (4) recognition of DNA molecules?

24 Which of the following properties are essential for the functioning of an aminoacyl synthetase: (1) recognition of a codon; (b) recognition of the anticodon of a tRNA molecule; (3) recognition of the amino-acid recognition loop of a tRNA molecule; (4) ability to distinguish one amino acid from another?

25 How can poly(U) be translated *in vitro* when it has no AUG and no Shine-Dalgarno sequence? If poly(A) were injected into the cytoplasm of an animal cell, would polylysine be made?

26 Which chain-termination codon could be formed by a single base change from UCG, UUG, and UAU?

27 In a eukaryote what would be the first three amino acids translated from the following mRNA—AGAUGAAUGAAUGCC...?

28 What is the minimum number of AUG sequences that would be in a tricistronic mRNA? Would it be likely that the actual number of AUGs would be larger?

29 An amber suppressor supplies an amino acid at the amber termination triplet UAG. It would seem as if, in a cell containing an amber suppressor, many necessary proteins might not be terminated, so that possession of this suppressor would be lethal. Why is this not the case?

ANSWERS TO DRILL QUESTIONS

1 A ribosome is a particle that binds mRNA and two tRNA molecules such that amino acids are aligned in the order that the codons are aligned in the mRNA.

2 $4^4 = 256$.

3 UUG, leucine; AAU, asparagine.

4 With G at the 3' terminus the poly(U) will terminate with UUG and the polypeptide will end with leucine; if it is at the 5' terminus, the poly(U) will start with GUU, so the polypeptide will initiate with glycine.

5 GCG corresponds to alanine, and CGC to arginine. Thus, the polypeptide would be ...AlaArgAlaArg....

6 Methionine and tryptophan.

7 UAA, UAG, UGA.

8 AUG; methionine.

9 Methionine and formylmethionine; in eukaryotes, only methionine.

10 In prokaryotes the start codon is the first AUG after (3' to) a Shine-Dalgarno sequence; in eukaryotes the start codon is the nearest AUG to the 5' end of the RNA.

11 In both prokaryotes and eukaryotes, two: $\mathrm{tRNA^{Met}}$ and the initiating tRNA.

12 The anticodon, DHU, and TψC loops

13 CCA-3'.

14 The 3' end.

15 It reacts with ATP such that AMP is linked via its phosphate to the COOH group of the amino acid.

16 Stem—a double-stranded region; loop—a closed single-stranded region attached to a stem; arm—a stem plus its attached loop; tertiary base pairs—base pairs, not necessarily between the bases themselves, causing the three-dimensional folding

of tRNA.

17 Base-base, base-ribose, and ribose-ribose bonds. The second and third type are used only for tertiary folding.

18 Hypoxanthine.

19 A, U, C.

20 A.

21 R'Q'P', because both codon and anticodon are by convention written with the 5' end at the left.

22 Leucine to tRNA^{Leu} and serine to seryl-tRNA^{Leu}.

23 1.

24 3 and 4.

25 In the *in vitro* system the high Mg concentration eliminates these requirements. No, because *in vivo* an AUG is required.

26 UAG.

27. Met, Asn, Glu.

28 The minimum is 3; the actual number would be larger since at least one of the proteins would contain at least one methionine.

29 UAG is usually preceded or followed by another chain termination codon.

PROBLEMS

1 Ribonuclease contains 124 amino acids. What is the least number of nucleotides you would expect to find in the gene encoding the protein? The first amino acid is not methionine.

2 Suppose that in another world the code is a doublet code having four bases. What is the maximum number of amino acids that could be in proteins?

3 At one time the possibility was considered that the genetic code might be an overlapping code of the following type: the codons in base sequence CATCATCAT... would be CAT ATC TCA CAT.... How was this hypothesis affected by the observation that mutant proteins usually differ from a wild-type protein by a single amino acid?

4 A portion of the coding strand of a DNA molecule has the sequence TTTTACGGGAATTAGAGTCGCAGGATC. What is the amino acid sequence of the polypeptide encoded in this region? Designate the amino end. Assume that the entire strand is transcribed and that a start codon is needed for initiation of polypeptide synthesis.

5 *E. coli* DNA has a molecular weight of 2.7×10^9. If all of this nucleotide sequence encoded proteins, how many proteins of average size would *E. coli* make? The molecular weights of an average amino acid, an average nucleotide pair, and an average protein are 110, 660, and 50,000, respectively.

6 When the repeating polynucleotide poly(UAUC) is used in a particular *in vitro* translation system, four different repeating polypeptides are formed—that is, poly(Try-Leu-Ser-Ile) with any of the four amino acids as the N-terminal amino acid. What does this result say about the choice of the first base read by this *in vitro* system?

7 When the repeating polynucleotide poly(AUAG) is used in an *in vitro* system that does not require AUG as a start codon, only two polypeptides are produced. What are they?

8 Suppose that you are making use of the alternating copolymer GUGU-GUGUGU... as an mRNA in an *in vitro* protein-synthesizing system. Assuming that an AUG start codon is not needed in the *in vitro* system, what peptides are made by this mRNA?

9 List the codons which, if altered by a single base change, could give rise to a chain termination codon.

10 The synthetic polymer poly(A) is used as an mRNA molecule in an *in vitro* protein-synthesizing system that does not need a special start codon. Polylysine is synthesized. A single guanine nucleotide is added to one end of the poly(A). Then, polylysine with a glutamic acid at the amino terminus is made. At what end of the poly(A) has the G been added, the 3' end or the 5' end?

11 What polypeptide products are made when the alternating polymer GUGU... is used in an *in vitro* protein-synthesizing system that does not need a start codon?

12 How would poly(AUAG) be used to show that UAG is a stop codon?

13 Two genes *A* and *B* are encoded in the same mRNA molecule. An added base pair in *A* eliminates the activity of gene *B* (no B protein is made). A second mutation in *A*, which is shown not to be either a base-pair addition or a base-pair deletion, restores activity of the B protein. What change was probably intro-

duced by the second mutation?

14 The central region of a polypeptide containing a single lysine has the amino acid sequence Phe-Leu-Tyr-Ala-Lys-Gly-Glu.... A mutation is found that causes the polypeptide to terminate with Phe-Leu-Tyr-Ala. Which of the lysine codons was used in synthesis of the protein?

15 Which of the following amino acid changes can result from a single base change? (1) Met→Arg, (2) His→Glu, (3) Gly→Ala, (4) Pro→Ala, (5) Tyr→Val.

16 Which amino acids can replace arginine by a single base-pair change?

17 Nitrous acid induces mutations by deaminating cytosine and adenine, thereby converting them respectively to uracil and hypoxanthine (which pairs like guanine). Which of the following amino acid changes could be induced by nitrous acid: Arg to His, Leu to Ser, Tyr to His, Met to Ile, Gly to Asp?

18 There are several arginine codons. Suppose you had a protein that contained only three arginines (Arg-l, Arg-2, Arg-3). In a particular mutant, Arg-1 is replaced by Gly. In another mutant, Arg-2 is replaced by Met. In still another mutant, Arg-3 is replaced by Ile. Suppose several hundred other mutants at various sites are isolated. Which other amino acids would you expect to find replacing Arg-1, Arg-2, and Arg-3, assuming only single-base changes?

19 Consider a segment of DNA in the very early part of a long protein. This small segment can tolerate many amino acid changes without altering the activity of the protein. The sequence is the following: TCGTAGAGGG|GCAATGAGC-AATACCCCGGA.... A mutation occurs in which two G's get inserted at the |. This change inactivates the protein, because the reading frame is altered by the two bases. However, if another mutation occurs that changes the underlined G to a C, an active protein (though not necessarily the same one) is made. Explain.

20 Two genes, *A* and *B*, are known from mapping experiments to be very near each other. A genetic deletion mutation is isolated that eliminates the activity of both genes *A* and *B*. In an attempt to study the nature of the deletion, all cellular proteins are isolated. Neither the A nor the B protein can be found, but a new protein is isolated that is larger than either of these proteins. The sequences of 30 amino acids at the amino and carboxyl termini of this are determined and found to be those at the amino terminus of the A protein and the carboxyl terminus of the B protein, respectively. (a) Are the two genes normally transcribed as part of a polycistronic mRNA molecule or as separate transcription units? (b) What can be said about the number of bases deleted? (A precise number cannot be stated, but a property of the number can.)

21 Explain what base changes will cause the following phenotype in yeast (a eukaryote). The amino terminus of a wild-type protein determined by a particular gene has the amino acid sequence Met-Leu-His-Tyr-Met-Gly-Asp-Tyr-Pro. A mutant (#1) is found that lacks activity of the gene, but chemical analysis indicates that the mutants contains a new protein, one having the sequence Met-Gly-Asp-Tyr-Pro at the amino terminus and the wild-type sequence at the carboxyl terminus. A second mutant (#2) is found that has also lost activity, but there is no sign of a mutant protein. A tripeptide not present in wild-type yeast is isolated in mutant #2, but you realize that it is not really necessary to determine its sequence. What changes in the base sequence have led to these mutants? What is the amino acid sequence of the tripeptide?

22 Consider two adjacent genes *A* and *B* in the same polycistronic mRNA. (a) An acridine-induced mutant that prevents synthesis of protein B maps in gene *A* rather than in gene *B*. Growth of this mutant in a medium containing an acridine yields A^+ revertants that are also B^+. Explain these events. (b) A B^+ revertant is isolated that also maps in gene *A*. This revertant is still A^-. Furthermore, growth in acridine can no longer induce formation of an A^+ mutant. If a suppressor is introduced into the B^+ revertant, the phenotype becomes A^-B^-. Explain what is happening.

23 Suppose you are synthesizing a random copolymer of G and U using equimolar quantities of G and U nucleotides. If this polymer is used as a mRNA, which amino acids can be incorporated into protein and at what relative frequencies?

24 For an *in vitro* protein synthesis experiment to study the genetic code, a synthetic mRNA species was synthesized with polynucleotide phosphorylase, using adenosine diphosphate (ADP) and cytidine diphosphate (CDP) in the ratio 5:l. The base sequence produced in this synthesis is random. The following amino acids were incorporated into polypeptide material in the following relative amounts: lysine, 100; asparagine, 20; glutamine, 20; threonine, 24; proline, 4.8; histidine, 4. Using this information, identify possible codons for the above amino acids and give their frequencies.

25 The first codon that was identified was UUU for phenylalanine. The insolubility of polyphenylalanine allowed several other codons to be identified by a simple trick. Each nucleotide X was separately coupled to the 3'-OH terminus of polyuridylic acid (poly(U)). The coupling was heterogeneous in that some poly(U) molecules terminated with one X moiety, some with two, some with three, etc. However, the addition of a single X was more probable than two, and two more probable than three, etc. Which amino acids will be at the terminus of polyphenylalanine when each of the four bases is separately coupled to poly(U)?

26 In starting their work on the nature of the genetic code, Crick and collaborators were confident that the N-terminal end of the T4 rIIB protein is not required for phage growth, i.e., it is, in that sense, dispensable. One of the reasons that they drew that conclusion was the existence of a deletion, named *r1589*, which spans

the N-terminal end of gene *rIIB* and the C-terminal third of gene *rIIA* and yields a fused rIIA-rIIB protein which has rIIB but not rIIA function. To judge from the genetic data and electron microscope-heteroduplex mapping, the *r1589* deletion is about 900 to 1100 nucleotides long. Following are four numbers that represent guesses regarding the exact length of the *r1589* deletion: 980, 984, 988, 1088. The base sequence of the DNA encoding the C-terminal end of the rIIA protein and the N-terminal end of the rIIB protein indicates that a 14-nucleotide spacer separates the genes. This information is sufficient to enable you to eliminate some of these proposed values. What values must be wrong? (To put it another way, what rule must be followed by the *r1589* deletion, and any other deletion that fuses the rIIA and rIIB polypeptide chains together?)

27 What three requirements must be met if an aminoacyl synthetase is to charge a tRNA molecule?

28 Protein synthesis occurs with fairly high fidelity. That is, it is very rare that a protein is made with incorrect amino acids. What features of the process of polypeptide synthesis are responsible for this fidelity?

29 What anticodons probably corresponds to the codon UGG?

30 What codons would be recognized by the anticodon UUC? Could glutamic acid have that codon?

31 Consider both the code as we know it and the wobble hypothesis, what is the minimum number of tRNA molecules needed to recognize the 61 codons corresponding to amino acids?

32 According to the wobble hypothesis the anticodon for the cysteine codons UGU and UGC could be either GCA or ICA. Consider the fact that UGA is a stop codon and deduce the cysteine anticodon.

33 In a segment of DNA containing overlapping genes would you expect the two proteins to terminate at the same site or have the same number of amino acids? Explain.

34 The codon UAG does not correspond to any amino acid. Some strains carry suppressors which are tRNA molecules mutated in the anticodon, enabling an amino acid to be placed in at a UAG site. Assuming that the anticodon of the suppressor differs by only one base from the original anticodon, which amino acids could be inserted at a UAG site? At a UAA site?

35 Consider two genes *a* and *b*. In growth medium Q, a mutation in either *a* or *b* is lethal (that is, the genotypes a^-b^+ and a^+b^- are lethal combinations. A genetic cross is made between one bacterium with genotype a^-b^+ and a second with genotype a^+b^-. It is found that there are two types of recombinants that can grow in Q, the a^+b^+ wild-type and (surprisingly) the a^-b^- double mutant—that is, a^- suppresses the b^- mutation and b^- suppresses the a^- mutation. Propose a molecular explanation for this phenomenon.

36 A mutant protein has amino acid B at a site normally occupied by A. If the strain also contains a missense suppressor, the protein has some activity. However it is known that the particular suppressor substitutes amino acid D for C and never substitutes A for B. Explain.

ANSWERS TO PROBLEMS

1 (3)(124) + 3 (termination) + 3 (initiation) = 378.

2 15. One needs one codon to separate coding sequences. It could be both a start and a stop.

3 In an overlapping code, a base change would affect three codons, so that often more than one amino acid would be changed. (One is possible because of the redundancy of the code.) Thus, an overlapping code was eliminated from consideration.

4 The start codon is complementary to the DNA base sequence TAC (bases 4, 5, 6). Note that an assumption must be made about the direction of reading the strand. The convention is to write the coding strand so that the mRNA can be read from left to right; that is, the coding strand is written with the 3' end at the left. The sequence is NH_2-Met-Pro-Leu-Ile-Ser-Ala-Ser.

5 If the average protein has a molecular weight of 50,000, it has about 454 amino acids, which corresponds to 3(454) = 1362 base pairs, or a DNA molecular weight of 8.25 million. Dividing the molecular weight of the bacterial DNA by this value, yields about 3000 genes, or 3000 polypeptides.

6 Any base can serve as the first base.

7 Remember that UAG is a stop codon. The only reading frames that can produce a polypeptide are AUA GAU AGA UAG, which yields the tripeptide Ile-Asp-Arg, and GAU AGA UAG, which yields the dipeptide Asp-Arg, are the only ones; AGA UAG... and UAG... cannot.

8 ...Val-Cys-Val-Cys-Val-Cys... and peptides of various sizes starting with Val or Cys.

9 AAA, AGA, CAA, CGA, GAA, GGA, UUA, UCA UAC, UAU, UGC, UGG,

UGU, UUG, UCG, CAG, GAG, AAG.

10 Codons are read from the 5' end of the mRNA, and the first amino acid in a polypeptide chain is at the amino terminus. If the G were at the 5' end, the first codon would be GAA (glutamic acid); if it were at the 3' end, the last codon would be AAG (arginine), and arginine would be at the carboxyl terminus. Thus, the G is at the 5' end.

11 GUG codes for valine and UGU codes for cysteine. Therefore, an alternating polypeptide containing only valine and cysteine would be made.

12 The repeating tetrapeptide can be read starting with either A, U, A, or G, in that order, yielding the codon sequences, AUA GAU AGA UAG (Ile Asp Arg Stop), UAG AUA GAU AGA (no protein made), AGA UAG (Arg only), and GAU AGA UAG (Asp Arg Stop), if UAG were a stop codon. The second and third sequences would not yield a polypeptide. Detection of either the tripeptide Ile-Asp-Arg or the dipeptide Asp-Arg indicate that UAG is a stop codon. If none of the codons were stop codons, long polypeptides would form. If either AUA, GAU, or AGA were stop codons, different polypeptides would be made.

13 The addition changes the reading frame of the first sequence so that the stop codon for the A protein is not encountered. Thus, the AUG of the B protein is not read. A change producing a stop codon between the start of A and the AUG for B will stop translation of the *A* gene and allow an in-phase restart at the AUG of B. Note that the stop codon can be either before or after the insertion.

14 There is no way of knowing because both codons could be altered to yield the stop codons UAA and UAG.

15 (1), (3), (4).

16 His, Glu, Cys, Trp, Ser, Gly, Leu, Pro, Ile, Thr, Lys, and Met.

17 Arg to His, Met to Ile, and Gly to Asp.

18 Arg-2 could be replaced by Gly, Trp, Lys, Thr, or Ser. Arg-3 could be replaced by Ser, Lys, Thr, and Gly. The Arg-1 codon cannot be identified.

19 The G-to-C mutation creates a UAG stop codon. Thus, the AUG just downstream and in the correct reading frame can be used to start in-frame synthesis again.

20 (a) The deletion has removed the terminus of one of the coding sequences and the early part of the other coding sequence in a polycistronic mRNA molecule. This is known as a gene fusion. (b) It must be a multiple of three. Otherwise, the carboxyl terminus of the protein would be in a different reading frame and

would not match the corresponding terminal sequence of the altered protein.

21 Mutant #1 could arise in two ways. It could have a deletion of 12 base pairs, those encoding amino acids 1–4 or 2–5. Alternatively, it could have a mutation in the start codon for the first Met. If that were the case, the second AUG would be used for starting. Note that the change in AUG might not work in a prokaryote, because the second AUG might be too far from the Shine-Dalgarno sequence. However, in a eukaryote the first AUG is the one used for starting. Mutation #2 is a change from Tyr (UAU or UAC) to Stop (UAA or UAG); the tripeptide is Met-Leu-His.

22 (a) A frameshift mutation has not only altered the protein but removed the stop codon of gene *A* from the reading frame. Thus, the AUG of *B* is not encountered. (b) A stop codon that is in the new reading frame in *A* has been introduced. This allows the AUG of gene *B* to start a new reading frame. The suppressor prevents chain-termination of the mutant *A*.

23 The frequencies of the triplets are: (3G), 1/8; (3U), 1/8; (2G,1U), 3/8; (2U,1G), 3/8. The (2G,1U) combination yields equal amounts of Gly (GGU), Val (GUG), and Trp (UGG). The (2U, 1G) combination yields equal amounts of Leu (UUG), Cys (UGU), and Val (GUU). GGG and UUU yield Gly and Phe respectively. The totals are (Val and Gly appears twice): Gly, 1/4; Trp, 1/8; Leu, 1/8; Cys, 1/8; Phe,1/8; Val, 1/4.

24 In any sequence of three bases in which A occurs five times as often as C, the probabilities of having A or C in a particular position are 5/6 and 1/6 respectively. Therefore, the probabilities for each triplet are: AAA, $(5/6)^3 =$ 125/216; ACA, AAC, CAA, 5/6 x 5/6 x 1/6 = 25/216 for each; ACC, CAC, CCA, (5/6)(1/6)(1/6) = 5/216 each; CCC, $(1/6)^3 = 1/216$. If the relative amount of AAA is defined to be 100, then the relative amounts of each of the other triplets are: ACA, AAC, CAA, (25/125)(100) = 20 each; ACC, CAC, CCA, (5/125)(100) = 4 each; (1/125)(100) = 0.8. Therefore, the codon assignments are: Lys, AAA; Asn, (2A,1C); Gln, (2A,1C); His, (1A,2C); Pro, CCC and (1A,2C); Thr, (2A,1C) and (1A,2C). When the bases are given in parentheses, their order in the codon is not known.

25 If A were added, the terminating codons would be UUA, UAA, and AAA, in decreasing frequency. The amino acids attached to polyphenylalanine would be leucine and lysine (UAA is a stop codon). For C and G the amino acids are serine (UCC) and proline (CCC), leucine (UUG), tryptophan, and glycine (GGG).

26 The rIIA and rIIB proteins are not translated in the same absolute reading frame—there are 14 nucleotides between the C-terminal Ala of the rIIA protein and the N-terminal fMet of the rIIB protein. Any deletion fusing rIIA and rIIB peptides must put them into the same frame. The deletion must take out $14 + 3n$ nucleotides, in which n is any integer. The values 984 – 14 = 970 and 988 – 14 =

974 are not multiples of 3 and must be wrong; the values 980 – 14 = 970 and 1088 – 14 = 1074 are multiples of 3 and could be right.

27 The amino acid must be present in an activated form (actually an amino acid-AMP complex), the synthetase must be able to bind to the tRNA, and the three-dimensional structure of the tRNA molecule must be such that the CCA (3') terminus is positioned in the active site of the synthetase.

28 Specificity of (1) codon-anticodon binding and (2) recognition of a tRNA molecule by the corresponding aminoacyl synthetase.

29 The base that could pair with the G in the third position could be C or U. If it were U, it would also pair with the codon UGA, which is a stop codon. Thus, it is not U and must be C. Therefore, the anticodon is 5'-CCA.

30 In the first codon position, U can pair with A or G. Thus, the codons are GAA and GAG. Yes, because glutamic acid has only two codons.

31 32.

32 If the anticodon were ICA, the codon UGA would also be a Cys codon, according to the wobble hypothesis. If GCA were the Cys anticodon, UGA would be reserved as a stop codon. Thus, the Cys anticodon is GCA.

33 They will not terminate at the same site because no sequence of four bases contains two overlapping stop codons. They would not usually have the same number of amino acids, but certainly could, depending on where each protein starts and stops.

34 For UAG, the amino acids are Tyr, Leu, Trp, Ser, Lys, Glu, and Gln. For UAA they are Tyr, Lys, Glu, Gln, Leu, and Ser.

35 The products of genes *a* and *b* mutually interact. The mutational changes in each are in the interaction sites and are compensatory.

36 Amino acid C normally interacts with A. Neither C and B nor A and D interact, but B and D can interact. Thus, C elsewhere in the protein is replaced by a D so that the protein can fold properly.

MORE CHALLENGING PROBLEMS

1 You have isolated an acridine-induced mutant of a particular protein and have purified the mutant protein. If the wild-type and mutant proteins are cleaved with trypsin and the peptides are examined, it is found that only one peptide is altered. (a) Where is this peptide located in the protein? (b) The peptide from the wild-type protein is Leu-Met-Ser-Val-Glu; that from the mutant is Leu-Thr-Glu-Arg. How many extra bases have been added to the DNA by growth in the presence of an acridine? Identify and locate the new base or bases in the base sequence.

2 Under what circumstances might the appearance of a chain termination codon not result in loss of activity of a protein?

3 Hemoglobin is synthesized from an mRNA molecule that has the 3'-terminal sequence AAGUAUCACUAAGC-3'. The carboxyl-terminal dipeptide of hemoglobin is ...Tyr-His-COOH. A number of mutant hemoglobins from patients with genetic diseases exhibit elongated polypeptide chains. One such mutant hemoglobin ends in ...Tyr-His-Leu-Ser-COOH. What is the mutation responsible for this elongated hemoglobin? Hint: Only one base is involved. There are no known nonsense suppressor mutations in human cells, so the mRNA for hemoglobin must itself be altered. Assume that protein synthesis can terminate at the 3' end of a molecule of mRNA as well as at a termination signal before the 3' end.

4 The diagram below represents a piece of DNA from the *middle* of a wild-type gene. Spaces between the bases indicate the reading frame of the corresponding mRNA. X and Y are unknown bases; a prime indicates a complementary base.

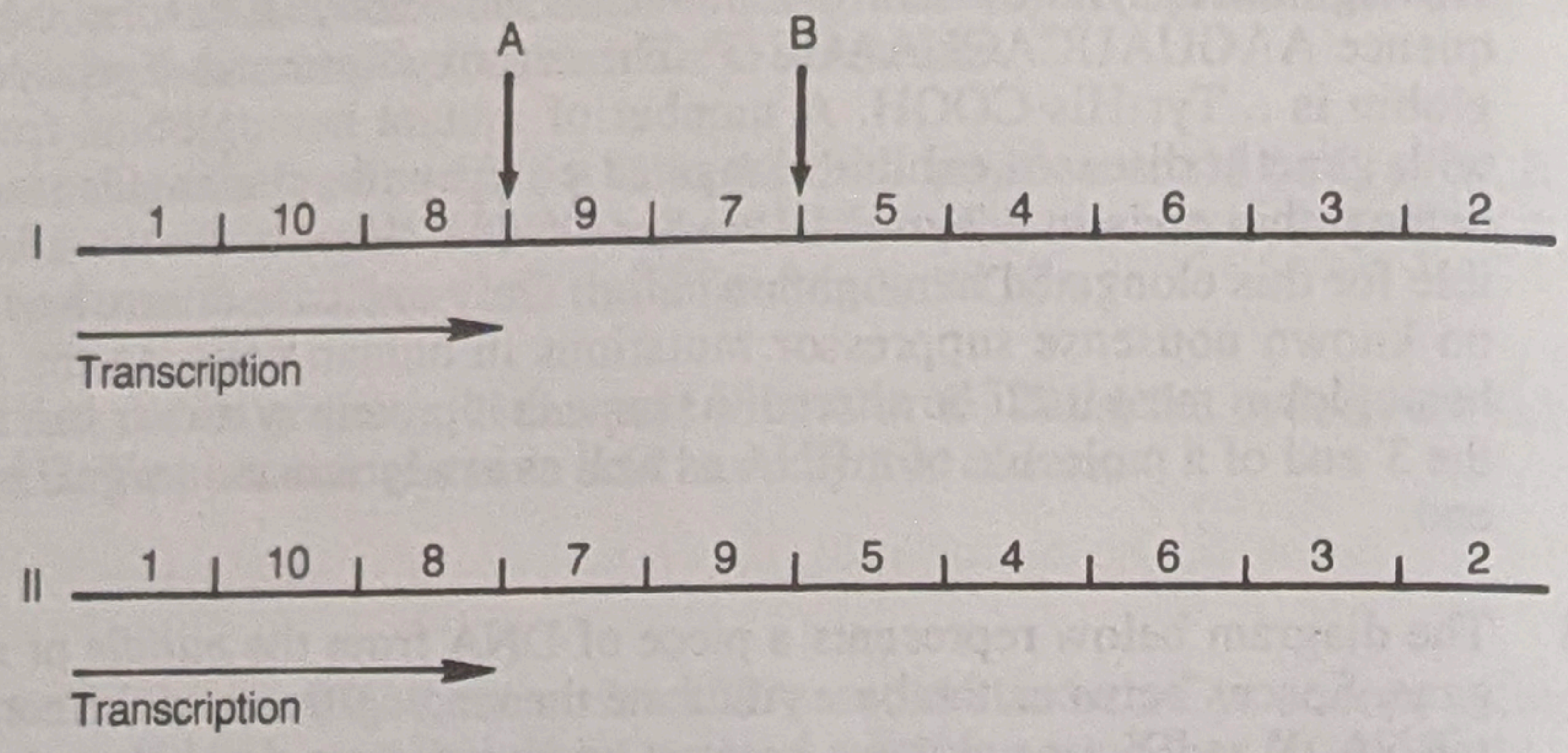
5' X ATA TAG GGG GCA Y 3' Strand 1
3' X' TAT ATC CCC CGT Y' 5' Strand 2

(a) What is the sequence of amino acids in the indicated four residues? Label the amino and the carboxyl termini. (b) In a certain mutant you find the amino acid sequence Cys-Pro-Tyr-Met. Assuming the occurrence of only a simple mutational event (that is, one not involving several base pairs), how does the mutant DNA sequence differ from that of the wild-type? What type of mutation is this? (c) What must base X be, and why?

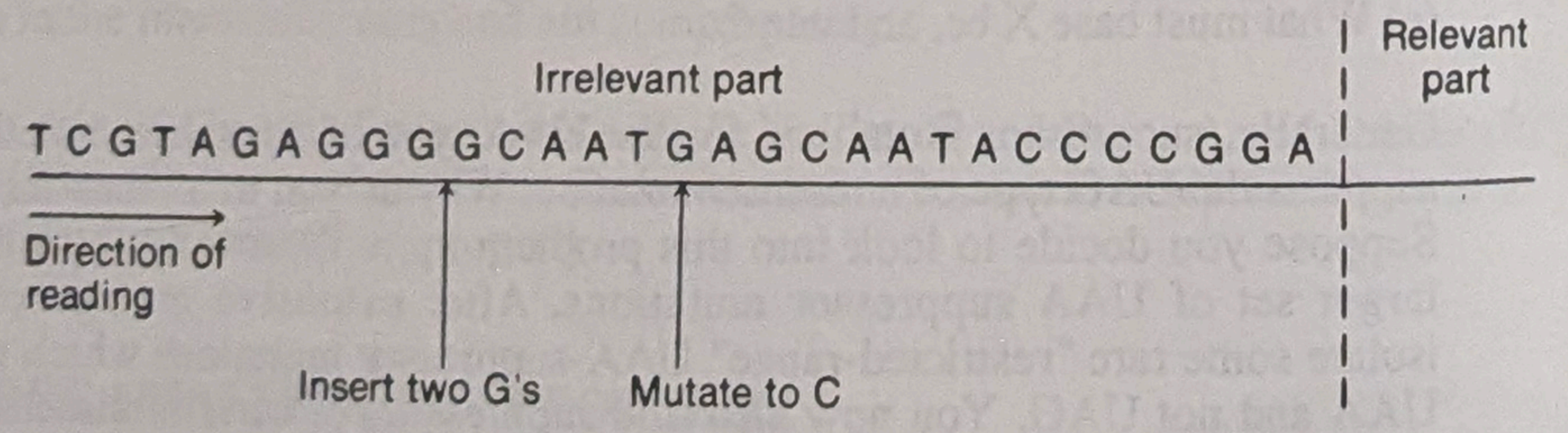
5 Generally, suppressor mutations for the UAA type of nonsense mutation also suppress the UAG type of nonsense mutation. Why do you think that this occurs? Suppose you decide to look into this problem more thoroughly by isolating a larger set of UAA suppressor mutations. After extensive mutagenesis, you isolate some rare "restricted-range" UAA-suppressor mutations which suppress UAA and not UAG. You now also find suppression of certain missense mutations. Propose an explanation in terms of tRNA structure. What amino acid do you predict will be replaced by the latter class of suppressor tRNA molecules?

6 The ten histidine-biosynthetic enzymes in *E. coli* form a single polycistronic mRNA, as shown below (I), where each number represents a gene. (The num-

bers also indicate the sequence of the encoded enzymes in the biosynthetic pathway. Suppose that you isolate a mutant that somehow has undergone a perfect inversion of two genes by breaking and rejoining at the points shown by arrows, in such a way that the entire nucleotide sequences of genes *5*, *7*, *8*, and *9* are preserved, but the order now is that shown in sequence II. (a) Would you expect this mutant to be His$^+$ (normal phenotype) or His$^-$? (b) Explain how, if at all, the inversion might affect the transcription of gene *4*. (c) Explain how, if at all, the inversion might affect the translation of gene *4* message.



7 Consider a piece of DNA in a short nonessential (i.e., any amino acid sequence will function) part of a gene, shown in the figure below. Note that only the strand from which the mRNA is copied is shown. (The other strand is irrelevant to this problem.) Suppose that, by acridine mutagenesis, a new sequence is created, so two G's are inserted at the point indicated. This will inactivate the protein because reading will be out of phase. However, if the G at the second indicated point is then mutated to become a C, an active protein (but not necessarily the same one) is made. Explain.



ANSWERS TO MORE CHALLENGING PROBLEMS

1 (a) It must be the C-terminal peptide. (b) The key to solving this problem is noting that the Met codon must be AUG and that shortening the peptide means that a chain termination codon has been generated. Since the terminal Glu codon in the wild-type sequence is GA_ and the Val codon is GU_, the simplest explanation for termination is that the Val codon is GUU and there is a single base insertion that joins the second U to the GA of the Glu codon. Since the Thr codons all start with AC and Met is gone, a C must be inserted to the right of the A in the Met codon.

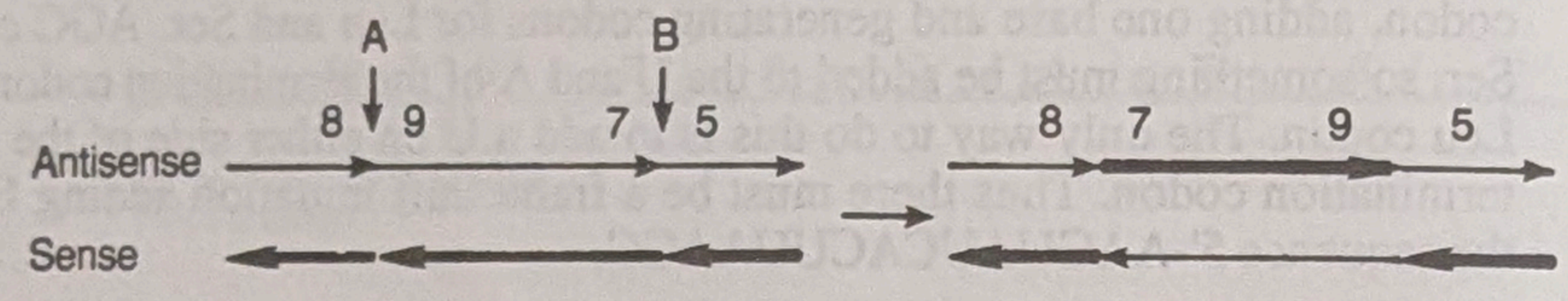
2 If it is very near the terminus of the protein so that only a few nonessential amino acids are removed.

3 The Tyr and His codons followed by a termination (UAA) codon are easy to find in the mRNA. To add two more amino acids, one has to add a nucleotide to make two extra codons. This must be done in such a way as to break up the termination codon, adding one base and generating codons for Leu and Ser. AGC codes for Ser, so something must be added to the U and A of the termination codon to get a Leu codon. The only way to do this is to add a U on either side of the U in the termination codon. Thus there must be a frameshift mutation adding U to give the sequence 5'-AAGUAUCACUUAAGC.

4 (a) The transcript from DNA strand 2 is 5'-X-AUA-UAG-GGG-GCA-Y-3'. The transcript from DNA strand 1 is 5'-Y'-UGU-CCC-CUA-UAU-X'-3'. The DNA strand 2 transcript contains a stop codon so that the coding strand is 1 and the amino acid sequence is Cys-Pro-Leu-Tyr-COOH. (b) The mRNA sequence for such an amino acid sequence is UG(U,C)-CC(U,C,G,A)-UA(U,C)-AUG, in which the symbols in parentheses represent alternative bases at one site, each of which would yield the same amino acid sequences. Given the base sequence of the wild type, the mutant sequence must be Y'-UGC-CCC-UAU-AUG, which cannot be generated from the original sequence by a single base substitution. Instead, a deletion of a C in position 3, 4, 5, 6, or 7 (starting with UGC) must have occurred. (c) From (b) and since the third base of the Met codon is G, X' must be G and X must be C.

5 Both types are suppressed because of wobble; the anticodon for UAA is UUA and the first anticodon U can, by wobble, pair with a G in the third position of the codon. Thus, the anticodon UUA can pair with the codon UAG. Since a UAA codon can also, by wobble, be suppressed by a tRNA with the IUA anticodon, which cannot pair with a UAG codon, such a mutant could account for the "restricted"-range suppressor. The amino acids is one whose tRNA molecule could be mutated to have an IUA anticodon. The amino acid must have a codon starting with UU, UC, UG, CA, GA, or UA, and be capable of having an I in the anticodon (that is, be at least triply redundant). Only serine (UCA, UCG, UCU, UCC) has this property.

6 (a) If the stretch of DNA corresponding to genes *9* and *7* is inverted and reinserted as described, the antisense strand of this segment must become connected to the sense strands of genes *8* and *5* to preserve polarity, as shown in the figure below. Since no new promoter is present in the mutant, the entire cluster will continue to be transcribed from left to right into a single mRNA, in which the segment corresponding to genes *7* and *9* will carry not the normal message, but its complement in reverse order instead. Therefore, the enzymes coded by genes *7* and *9* will not be produced, and the mutant will be His⁻. (b) The reversed antisense strand of the *9+7* region, which is not normally transcribed, may by chance contain a sequence that signals chain termination for mRNA synthesis. If so, gene *4* would not be transcribed. (c) Regardless of how it is spliced into the mRNA molecule, the reversed complementary message transcribed from the *9+7* segment is almost certain to include, by chance, one or more of the termination codons UAG, UAA, or UGA. The resulting polypeptide-chain termination will reduce the production of all the distal proteins, that is 5, 4, 6, 3, and 2.



7 The G-to-C mutation creates a UAG stop codon. Thus the AUG found "downstream" (and in the proper reading frame) can start in-frame synthesis going again. Although not shown, there must be a ribosome binding site somewhere to the left of the sequence.

CHAPTER 14

TRANSLATION II: THE MACHINERY AND THE CHEMICAL NATURE OF PROTEIN SYNTHESIS

CHAPTER SUMMARY

Ribosomes are ribonucleoprotein particles on which protein synthesis occurs. The ribosomes of prokaryotes are smaller than those of eukaryotes; their sizes are 70S and 80S respectively. Both prokaryotic and eukaryotic ribosomes consist of two subunits, one larger than the other. For prokaryotes the subunit sizes are 30S and 50S. The 30S subunit contains one 16S rRNA molecule and 21 different proteins. The 50S subunit contains one 5S rRNA and one 23S rRNA molecule and 31 different proteins. Thus, prokaryotic ribosomes contain three rRNA molecules and 52 different proteins. For eukaryotes the subunit sizes are 40S and 60S. The 40S subunit contains one 18S rRNA molecule and about 30 proteins. The 60S subunit contains one 5S, one 5.8S and one 28S rRNA molecule and about 50 proteins. Thus, the eukaryotic ribosome consists of four rRNA molecules and about 80 proteins. The ribosomes of chloroplasts and mitochondria have a prokaryotic rather than a eukaryotic structure. Ribosomes require the Mg^{2+} ion for stability; when the Mg^{2+} concentration is reduced, the particles break down successively into the two subunits and then each subunit dissociates, releasing the RNA and protein molecules. Under appropriate conditions ribosomes can be reconstituted.

In polypeptide synthesis mRNA is read in the 5'-to-3' direction. Polypeptides are synthesized beginning with the amino end. The participants in polypeptide synthesis are a ribosome, mRNA, charged tRNA molecules, GTP, and a variety of protein factors. The protein factors are needed for specific steps in chain initiation, elongation, and termination. Each ribosome contains two tRNA-binding sites, called the A site and the P site. In prokaryotes polypeptide synthesis begins with formation of the 30S preinitiation complex, which consists of mRNA, a 30S subunit, fMet-$tRNA^{fMet}$. Formation of the complex requires three initiation factors and cleavage of GTP. A 50S subunit binds to the preinitiation complex, forming the 70S initiation complex. The fMet-$tRNA^{fMet}$ occupies the P tRNA-binding site and its anticodon is hydrogen-bonded to the mRNA AUG codon at which synthesis is to begin. The protein factor EF-Tu delivers a second charged tRNA to the A site, the tRNA being selected by its ability to bind to the codon on the 3' side of the initiating AUG. The peptidyl transferase center of the ribosome, an active enzymatic site formed from portions of many ribosomal proteins, then forms a peptide bond between fMet and the amino acid carried by the tRNA in the A site. At the same time the fMet is cleaved from tRNA$^{\text{fMet}}$ and the tRNA leaves the P site.

The elongation factor EF-G then causes the amino acid in the A site to move to the P site, leaving the A site available for a second charged amino acid, which will be brought in by EF-Tu and which is again selected by codon-anticodon binding. The synthetic cycle consists of formation of a peptide bond between the amino acids bound to the tRNAs in the A and P sites, cleavage of the tRNA-amino acid bond of the tRNA in the P site, departure of the tRNA from the P site, and translocation of the tRNA in the A site to the P site. The cycle continues until a stop codon reaches the A site. Since no tRNA is available that can pair with a stop codon, the A site remains unoccupied. A release factor then alters the activity of the peptidyl transferase center, causing it to hydrolyze the bond holding the polypeptide chain to the tRNA in the P site. The polypeptide chain is released, the tRNA leaves the P site, and the 70S ribosome dissociates, ready to start protein synthesis again.

ATP and GTP, both of which contain high-energy phosphates, play different roles in protein synthesis. ATP is cleaved and its free energy of hydrolysis is used to form an activated amino acid that can be linked to tRNA; whereas GTP is an allosteric effector, binding to protein factors and causing conformational changes. In the overall process of protein synthesis ATP is used to form the peptide bond, and GTP is used in binding reactions and for translocation.

Proteins are often modified after synthesis. One modification that always occurs is removal of the formyl group from fMet. In some cases, the entire amino acid is removed. Other modifications are formation of disulfide bonds, cleavage of internal peptide bonds, hydroxylation of lysine and proline, acetylation of amino-terminal amino acids, glycosylation, and addition of lipids.

Once a ribosome has begun to move along a mRNA molecule, the initiating AUG codon becomes available for binding a second ribosome, both in prokaryotes and in eukaryotes. The result is that an mRNA molecule is typically bound to a series of ribosomes, each separated by about 80 nucleotides of mRNA. This structure is called a polysome. In prokaryotes, since RNA is synthesized in the 5'-to-3' direction, the Shine-Dalgarno sequence and the initiating AUG become available for ribosome binding before synthesis of the mRNA is completed. Thus, a ribosome typically binds to the nascent RNA. This coupled process does not take place in eukaryotes, because transcription occurs in the nucleus and translation in the cytoplasm.

A variety of antibiotics inhibit protein synthesis in prokaryotes. Most interfere with only one step. For example, puromycin, an antibiotic that is useful in research, is an analogue of charged tRNA that cannot form a peptide bond; hence, it causes premature termination. Chloramphenicol, another useful antibiotic, inhibits the peptidyl transferase center. Tetracycline, which has great clinical use, inhibits binding of charged tRNA to a 30S subunit. Streptomycin has a variety of effects on protein synthesis; most significant is inhibition of binding of tRNA$^{\text{fMet}}$ to the P site. Many antibiotics that are active in prokaryotes also inhibit protein synthesis in eukaryotes; this class has little or no clinical value because of their toxic effects on humans. Certain inhibitors are specific for eukaryotes—for example, abrin, cycloheximide,and ricin. Cycloheximide is useful as a cancer chemotherapeutic agent. Abrin and ricin are far too toxic for this use.

Protein synthesis in eukaryotes follows essentially the same scheme as that in prokaryotes. The principal differences are the greater number of protein factors required for many stages and the use of an initiator tRNA that is charged with Met rather than fMet. As mentioned earlier, the initiating AUG codon is selected by proximity to the 5' end of the mRNA rather than by a specific binding sequence, as is the case with prokaryotes. Eukaryotes make a variety of proteins that are secreted into vesicles and cisternae of a highly convoluted membrane system called the endoplasmic reticulum. These proteins have highly hydrophobic amino-terminal regions (a signal peptide), which causes the polypeptide chain to penetrate pores in the membrane system. The result is that many ribosomes are membrane-bound. The membranes plus the bound ribosomes constitute the rough endoplasmic reticulum. Penetration is not a random process but is caused by the action of a variety of recognition proteins and binding sites. The signal peptide is cleaved from the chain inside the cisterna, usually before the chain is finished. When synthesis of these proteins is completed, the ribosome is released from the membrane. Thus, there is no essential difference between membrane-bound and free ribosomes.

BOLD TERMS

A site, aminoacyl-tRNA, aminopeptidase, antibiotics, chain elongation, chain termination, chloramphenicol, chloroplasts, cisterna, coupled transcription-translation, cross-link, cycloheximide, deformylase, docking protein, EF-G, EF-Tu, elongation factor, endoplasmic reticulum, 50S subunit, 40S subunit, initiation factors, initiating tRNA hydrolase, P site, peptidyl site, peptidyl transferase, peptidyl transferase center, polypeptide chain initiation, polyprotein, polyribosome, polysome, puromycin, ribosome receptor, 70S ribosome, signal hypothesis, signal peptidase, signal recognition particle (SRP), SRP receptor, 30S preinitiation complex, 30S subunit, translocation, tRNA deacylase

ADDITIONAL INFORMATION

Several procedures for localization of proteins in ribosome subunits were described in the text. An additional useful technique is immuno-electron microscopy, which can be to locate proteins on the surface of a particle. In these experiments an antibody is first prepared by injecting a purified ribosomal protein into a rabbit. The antibody, which is a large protein, reacts only with the specific protein that elicited its production, and an antibody can be prepared for each purified protein. An antibody directed against a single ribosomal protein is mixed with either a 30S or 50S particle and the sample is then viewed by electron microscopy. An important feature of antibody molecules is that they have two identical binding sites, as discussed in Chapter 6. Thus, the antibody binds to its target protein on the surface of two identical ribosomal subunits, thereby forming a dimer. The point of contact of the two monomers is the same for both monomers and is the location of the particular protein.

Since the orientation of the 30S ribosome can be determined (as in Figure 14-4 of the text), the position of the particular protein can be localized by looking at particles in various orientations. These procedures have been repeated with antibodies directed against many ribosomal proteins.

Eukaryotes contain a variety of small morphologically distinct particles called organelles, which are responsible for carrying out specific functions. These organelles contain DNA and ribosomes, among other things. Two organelles, the mitochondrion (which contains components of the energy-generating system) and the chloroplast (which carries out photosynthesis) have well-characterized, protein-synthesizing systems. The most striking properties of these systems is that their ribosomes resemble those in prokaryotes both in size (near 70S) and the number and size of their RNA molecules; for example, their ribosomes have three rRNA molecules, as in prokaryotes, rather than four, which is characteristic of eukaryotes, and the sizes are 5S, 16S, and 23S, as in prokaryotes. Furthermore, many of the individual steps in protein synthesis are the same as those of prokaryotes. Evolutionists have hypothesized that these organelles may be the remnants of ancient prokaryotes that were parasitic and became established in eukaryotic cells.

DRILL QUESTIONS

1 How many nucleoprotein subunits are there in a prokaryotic and a eukaryotic ribosome?

2 How many RNA molecules are in a prokaryotic and a eukaryotic ribosome?

3 What are the *s* values of the ribosomes and subunits of prokaryotes and eukaryotes?

4 What are the s values of the RNA molecules of prokaryotic and eukaryotic ribosomes?

5 What are the L and S proteins of ribosomes?

6 Roughly how many different proteins are present in prokaryotic and eukaryotic ribosomes?

7 Which ribosomal subunit initially binds mRNA in prokaryotes and eukaryotes?

8 Do ribosomes have to be assembled with some special machinery or do they assemble spontaneously?

9 Do free (nontranslating) 70S ribosomes exist in bacteria?

10 With what polarity is mRNA read?

11 What nucleotide sequences in the mRNA signal the beginning and end of protein chains?

12 What is the direction of synthesis of a polypeptide chain?

13 Which amino acid is present at the amino terminus of an unprocessed polypeptide chain, in prokaryotes and eukaryotes?

14 What are the components of the 30S preinitiation complex in prokaryotes?

15 At what stage of polypeptide synthesis do 70S ribosomes form?

16 Indicate the amino and carboxyl amino acid of the polypeptide chain formed from the synthetic polymer poly(C) bearing a 3' A.

17 Which site, A or P, is initially occupied by the initiating tRNA molecule?

18 Which site, A or P, can never be occupied unless the other site is also occupied?

19 On which ribosomal subunit is the major portion of the A site?

20 Which of the following are steps in protein synthesis in prokaryotes: (1) binding of tRNA to a 30S particle; (2) binding of tRNA to a 70S ribosome; (3) coupling of an amino acid to a ribosome by an aminoacyl synthetase; (4) separation of the 70S ribosome to form 30S and 50S particles?

21 Which of the following statements are true of tRNA molecules: (1) they are needed because amino acids cannot stick to mRNA; (2) they are much smaller than mRNA molecules; (3) they are synthesized without the need for intermediary mRNA; (4) they bind amino acids without the need of any enzyme; (5) they occasionally recognize a stop codon if the Rho factor is present?

22 Why can peptidyl transferase not be purified, even in inactive form, free of ribosomes?

23 How many amino acids are linked together at the time the ribosome and mRNA first move with respect to one another?

24 What movement is probably responsible for the relative movement of mRNA and the ribosome (translocation)?

25 What protein factor participates in translocation?

26 Which of the following is the normal cause of chain termination: (1) the tRNA corresponding to a chain-termination triplet cannot bind an amino acid; (2) there is no tRNA with an anticodon corresponding to a chain termination triplet; (3) mRNA synthesis stops at a chain termination codon?

27 What enzyme is responsible for cleavage of the completed polypeptide chain from the last tRNA molecule?

28 Name three mechanisms that account for the lack of fMet in completed proteins.

29 Can more than one ribosome read a single mRNA molecule concurrently?

30 Why are transcription and translation not coupled in eukaryotes?

31 *In vivo* is heterogeneous nuclear RNA ever translated?

ANSWERS TO DRILL QUESTIONS

1 Two, in each case.

2 Prokaryotic, three; eukaryotic, four.

3 Prokaryotes—70S ribosome, 30S and 50S subunits; eukaryotes—80S ribosome, 40S and 60S subunits.

4 Prokaryotic—5S, 16S, 23S; eukaryotic—5S, 5.8S, 18S, 28S.

5 The L proteins are components of the large (50S or 60S) subunit; the S proteins are components of the small (30S or 40S) subunit.

6 Prokaryotic: 52; eukaryotic, about 80.

7 Prokaryotes, 30S; eukaryotes, 40S.

8 Spontaneously.

9 No. 70S ribosomes dissociate when translation is completed and do not re-form until the 30S subunit has bound mRNA, fMet-tRNA, and several initiation factors.

10 The 5′ end is read first.

11 Start—AUG. Stop—UAA, UAG, UGA.

12 The amino end is made first.

13 Prokaryotes—*N*-formylmethionine; eukaryotes—methionine.

14 A 30S ribosome, mRNA, $\text{fMet-tRNA}^{\text{fMet}}$, and initiation factors.

15 The 70S ribosome form when the 50S subunit joins the 30S preinitiation complex.

16 CCC corresponds to leucine and AAC to histidine. Leucine is at the amino end; histidine is at the carboxyl end.

17 The P site.

18 The A site.

19 50S subunit.

20 Steps 1, 2, and 4.

21 (1), (2), (3).

22 It is formed from parts of ribosomal proteins.

23 Two, those in the A and P sites.

24 The movement of peptidyl-tRNA from the A site to the P site.

25 EF-G.

26 (2).

27 Peptidyl transferase. Under the influence of a release factor, it becomes a hydrolyzing enzyme and cleaves the RNA-protein bond.

28 Removal of the formyl group, removal of fMet, and removal of a terminal oligopeptide.

29 Yes, and a polysome results when that occurs.

30 Transcription occurs in the nucleus, the location of the DNA and RNA poly-

merase. Translation occurs in the cytoplasm.

31 It is in the nucleus and translation occurs in the cytoplasm. It is believed that transport of RNA to the cytoplasm requires RNA processing.

PROBLEMS

1 Excluding the enzymes needed to synthesize the amino acids, make a rough count of the number of known enzymes required to synthesize a single protein molecule. What fraction of *E. coli* enzymes is involved in protein synthesis?

2 The synthetic polynucleotide 5'-AGGUUAUAGGAAAAA-3' is used as an mRNA in a system that does not require a start codon. What polypeptide is synthesized? Indicate the amino and carboxyl termini.

3 What physical feature do tRNA and rRNA have in common?

4 Suppose you reconstitute a ribosome, omitting a particular ribosomal protein, and observe that tRNA will not bind to the reconstituted particle. Can you conclude that tRNA binds to that protein?

5 Which of the following are roles of tRNA in prokaryotes: (1) to attach the amino acids to one another; (2) to bring the amino acids to the correct position with respect to one another; (3) to increase the effective concentration of amino acids; (4) to attach the mRNA to the ribosome?

6 How do the roles of ATP and GTP differ in polypeptide synthesis?

7 Translation has evolved in a particular polarity with respect to the mRNA molecule. What would be the disadvantages of having the reverse polarity?

8 A batch of reticulocytes actively making the protein hemoglobin were given radioactive amino acids for a brief period (where a "brief period" is a time that is short compared to the time required to polymerize the amino acids of a protein chain). Immediately thereafter, the soluble hemoglobin (that which is not attached to ribosomes) is examined for distribution of radioactivity in the chain. Which one of the following will be observed? (1) The radioactive label is found in amino acids at the amino end only. (2) Label is found in amino acids at all positions. (3) Label is found at the carboxyl end only.

9 Refer to Problem 8. Suppose the reticulocytes were permitted to continue making hemoglobin with nonradioactive amino acids for a long time following the radioactive pulse. If the soluble hemoglobin were examined, which one of the following would be found? (1) Label at the amino end only. (2) Label in amino acids at all positions. (3) Label at the carboxyl end only.

10 What are the requirements for the union of one 30S and one 50S ribosomal subunit *in vivo* and *in vitro*?

11 Which of the following is the normal cause of chain termination: (1) the tRNA corresponding to a chain-termination triplet cannot bind an amino acid; (2) there is no tRNA with an anticodon corresponding to a chain-termination triplet; (3) mRNA synthesis stops at a chain termination triplet?

12 In an *in vitro* protein-synthesizing system that requires an AUG in the mRNA, all proteins start with fMet. Consider a mRNA that has the sequence 5'-AUGAAAAAAAA. (a) What is the amino acid sequence of the protein that is made? (b) If an antibiotic that inhibits translocation is added, what will be the product, if any, of the *in vitro* synthesis?

13 Which processes in protein synthesis require hydrolysis of GTP?

14 Explain the function of initiation, elongation, and termination factors.

15 How many high-energy phosphate-containing molecules are required to translate the mRNA of a protein molecule containing 100 amino acids? Assume that the protein-synthesizing system contains free amino acids, ribosomes, all necessary enzymes and factors, ATP, and GTP.

16 Suppose you have isolated a temperature-sensitive mutant of *E. coli*. At 42 C it is defective in protein synthesis. A culture growing at 30 C is warmed to 42 C at the same time ^{14}C-leucine is added. After 10 minutes at 42 C trichloracetic acid is added (to precipitate macromolecules) and the acid-insoluble material is isolated and examined; it is found to be very radioactive. If instead the cells are broken open and the cell extract is fractionated by centrifugation, it is found that all ^{14}C is in a fraction containing ribosomes; there is no free ^{14}C-labeled protein in this case. What process is probably temperature-sensitive?

17 Suppose you have a mutant of *E. coli* which grows normally at 30 C but slowly at 42 C. There is nothing obviously defective about the rates of DNA and RNA formation or protein synthesis shortly after the temperature is increased from 30 C to 42 C. In studying the induction of the synthesis of a particular enzyme with this mutant it is observed that if the inducer is removed, the rate of synthesis of the enzyme at 30 C decreases with a half-life of 5 minutes. At 42 C, however, the rate of synthesis does not decline but the amount synthesized per generation per cell decreases with a half-life of one generation. What process is probably inhibited at 42 C?

18 How does peptidyl transferase differ from a typical enzyme?

19 Apart from the difference in the number of factors needed for protein synthesis

in prokaryotes and eukaryotes, what are the major differences in the modes of initiation?

20 A poly(A) sample terminated with G is used in an *in vitro* translation system. The molecules are heterogeneous with respect to length, as is normally the case. Synthesis is allowed to proceed for quite a long time, so at the time the reaction is terminated the fraction of molecules that are unfinished is exceedingly small. Clearly a major product is polylysine (AAA = lysine) terminated with serine (AAG). What other product is also made in large quantities?

21 A polysome has a 3' and a 5' end. At which end will the polypeptide chain attached to the ribosome be longer?

22 Suppose polyribosomes did not exist and only a single ribosome could bind to a mRNA at any moment. Compared to a system having polyribosomes, will the number of protein molecules synthesized per unit time be greater, less, or the same? Explain.

23 What is the relation between the rate of codon production (in mRNA synthesis) and the rate of amino acid polymerization?

24 How long does it take to synthesize a typical protein (molecular weight 50,000) at 37 C?

25 A chemical compound is found that, when added to growing cells, immediately inhibits the dissociation of a 70S prokaryotic ribosome when polypeptide synthesis is complete. It is found that in the course of time the compound also prevents initiation of protein synthesis. Why?

26 Penicillin, tetracycline, erythromycin, chloromycetin, and puromycin are powerful inhibitors of bacterial growth. However, only the first three of these are used in the treatment of human diseases. Chloromycetin is used only in cases where all else fails. Explain these facts in terms of microbial and mammalian protein synthesis.

ANSWERS TO PROBLEMS

1 RNA polymerase to transcribe mRNA and tRNA, a few enzymes for processing tRNA, about 20 tRNA synthetases, several enzymes for regenerating GTP, one or a few for forming a peptide bond, one or a few for releasing a finished polypeptide and a few for protein processing—perhaps 40 to 50 enzymes in all. This represents about 2% of the *E. coli* enzymes.

2 NH_2-Arg-Leu-COOH, when reading starts exactly at the 5' terminus. Reading

can begin at other positions also yielding NH_2-Gly-Tyr-Arg-Lys-COOH and NH_2-Val-Ile-Gly-Lys-COOH and fragments thereof.

3 They have extensive double-stranded regions.

4 No. The protein may be required for the overall three-dimensional shape of the ribosome.

5 (2), and possibly (3) and (4).

6 ATP is used to activate amino acids and hence indirectly provides the free energy for forming a peptide bond. GTP induces conformational changes by binding to proteins; the original conformation is restored when GTP is hydrolyzed. GTP is also used in an unknown way in translocation.

7 Synthesis would have to await the completion of a molecule of mRNA. In the existing system protein synthesis can occur while the mRNA is being copied from the DNA. Thus, protein synthesis can start earlier than would be possible with reverse polarity, and the mRNA is relatively resistant to nuclease attack.

8 (3) Label is found at the carboxyl end only. This experiment was originally used to determine the polarity of protein synthesis.

9 (2).

10 Formylmethionyl-tRNA and mRNA plus (perhaps) other things are required *in vivo;* a high Mg^{2+} concentration is required *in vitro.*

11 (2).

12 (a) Met Lys Lys Lys.... (b) Met Lys, because both P and A sites can be filled without the occurrence of translocation.

13 Formation of the 70S initiation complex and translocation.

14 Initiation factors are used to prevent interaction of the ribosomal subunits and for binding of tRNA to its ribosomal binding site; elongation factors are used to move tRNA from one ribosomal binding site to the next; termination factors are needed for release of the finished polypeptide chains and of the mRNA from the ribosome.

15 399 molecules: 200 to activate 100 amino acids, 198 to form 99 peptide bonds and 1 for initiation.

16 Release of finished proteins from the ribosome.

17 Degradation of mRNA.

18 It is not a single protein but is a catalytically active site that is formed from portions of many proteins.

19 In eukaryotes the initiating amino acid is Met rather than fMet and there is no Shine-Dalgarno sequence; instead the AUG nearest the 5' terminus of the mRNA is the initiating codon.

20 Polylysine terminated with serine is made only when translation begins at an AAA that is separated from the 3' terminus by 3*n* nucleotides. If reading begins at another position, the 3' G will not be read and the polypeptide will simply be polylysine.

21 The 5' end, since ribosomes move along the mRNA from the 5' end to the 3' end.

22 Less, because with polyribosomes several protein molecules can be synthesized simultaneously from a single mRNA molecule.

23 They are nearly the same: 17 amino acids polymerized per second and 55 nucleotides or 18 codons synthesized per second.

24 Such a protein has 5000/110 = 455 amino acids so it is synthesized in 455/17 = 27 seconds.

25 Protein synthesis cannot start with a 70S ribosome, and soon no 30S particles will be available for forming the preinitiation complex.

26 Penicillin interferes with bacterial cell wall synthesis and therefore has no effect on eukaryotes. Actinomycin affects transcription in both bacteria and mammals and hence is not of clinical usefulness. The remaining antibiotics affect translation, but tetracycline and erythromycin cannot inhibit translation in mammalian cells. Thus, these two drugs are useful against infection. Puromycin and chloromycetin are active against both bacterial and mammalian ribosomes. However, in cases in which an infection is becoming dangerous to a patient, chloromycetin can be used but the toxic effects have to be dealt with in other ways.

MORE CHALLENGING PROBLEMS

1 A melting curve (optical absorbance *vs.* temperature) can be obtained for RNA just as for DNA. The melting curve for a mixture of purified 16S and 23S rRNA is virtually identical to that for 70S ribosomes (as long as one corrects the observed value of the absorbance for light scattering). What can you conclude from this observation?

2 Suppose you are studying protein synthesis *in vitro* with the synthetic polymer AUGUGUUAG. With your standard preparation of factors and ribosomes you find the expected production of free fMet-Cys. If you make the same standard factor preparation after infection by phage X19, you find no free dipeptide, but do find Val-Leu bound to the ribosome. Propose an explanation for this observation.

3 *In vitro* protein synthesis is being carried out using $(AUG)U_{90}G_{90}$ as an mRNA molecule. The system has been obtained from bacteria known to carry a temperature-sensitive mutation that blocks protein synthesis *in vivo* within a few seconds after the temperature is raised. *In vitro,* if the temperature is raised a few seconds after initiation, the longest polypeptide found is Met-Phe_{30}, though some with fewer phenylalanines are occasionally observed. Suggest a defect in this mutant system.

4 Suppose the mRNA in Challenging Problem 3 had not started with AUG, and the explanation for the defect had been the same. How would the experimental result have been different?

ANSWERS TO CHALLENGING PROBLEMS

1 The RNA in the ribosome is extensively double-stranded, as is the free rRNA.

2 An initiator tRNA responding to the GUG codon has been made in the infected cell. Possibly the tRNA responding to AUG has been inactivated also. There is no termination codon in this reading frame so the peptide is not released.

3 There seems to be nothing wrong with the protein synthetic system itself. However, the system is unable to translate from the G portion of the mRNA. The simplest explanation is that no charged tRNA is present that can read the GGG codon. This could result from a temperature-sensitive tRNA molecule or aminoacyl synthetase.

4 The system would have to be one that did not need an initial AUG. In that case, translation would not always start at the end of the mRNA. This would have two consequences. First, the polypeptide would not start with Met, but this is obvious. Second and more important, some of the polypeptides would contain fewer than 30 phenylalanines, when translation began within the mRNA. Third, the reading frame would sometimes be different, so the final codon would sometimes be either UUG or UGG. Therefore, some of the polypeptides would have terminated with leucine or tryptophan.

CHAPTER 15

REGULATION OF THE ACTIVITY OF GENES AND OF GENE PRODUCTS IN PROKARYOTES

CHAPTER SUMMARY

Gene activity is regulated in prokaryotes in order to avoid waste and thereby to maximize the growth rate. The rule is simply that genes are turned on when expression is needed and off when not needed. Regulation is necessary because if all genes were expressed at all times, there would be enormous waste of amino acids, ATP, and other essential molecules. Regulation of gene activity usually occurs by control of transcription, but translational regulation occurs in a few cases. Transcriptional regulation can occur at the level of initiation or of premature termination. Initiation can be controlled by either negative or positive regulation. In negative regulation an inhibitor of transcription is present in the cell and the inhibitor is inactivated when gene expression is required. The anti-inhibitor is called an inducer. In positive regulation an effector molecule activates a promoter. In both types of regulation a protein is present that binds to the promoter. Some systems are both negatively and positively regulated.

The products of some genes are required continuously, for example, the enzymes required to metabolize glucose. To avoid excessive and wasteful synthesis of the enzymes, transcription is often regulated by the concentration of the enzyme itself; this is called autoregulation. This can be accomplished in several ways; a common mechanism is direct binding of the enzyme to a regulatory site in or near the promoter for the gene. When the sequence is occupied, transcription does not occur. Thus, if the concentration of the enzyme were to fluctuate for any reason, the rate of transcription would also fluctuate, but in a way that would restore the enzyme concentration to the required value.

The synthesis of degradative enzymes needed only on occasion, such as the enzymes required to metabolize lactose, a sugar encountered only infrequently in the environment, are regulated by an off-on mechanism. When lactose is present, transcription of the genes encoding the enzymes required to metabolize lactose occurs; when lactose is absent, such transcription does not take place. Lactose gene expression is negatively regulated. Two enzymes required to degrade lactose, permease (required for entry of lactose into bacteria) and β-galactosidase (the actual degrading enzyme), are encoded in a single polycistronic mRNA molecule, *lac* mRNA. Immediately adjacent to the *lac* promoter is a regulatory sequence of bases, the operator. A protein inhibitor, the *lac* repressor, is made by still another adjacent gene, and this protein binds tightly to the operator, thereby preventing RNA polymerase from initiating transcription at the promoter. Lactose is an inducer of transcription, because it can bind to the repressor, thereby preventing the repressor from interacting with the operator. Therefore, in the presence of lactose, there is no active repressor, the operator is unoccupied, and the *lac* promoter is always available to RNA polymerase. The repressor gene, the operator, the promoter, and the structural genes are adjacent to one another (separated only by small spacers, except between the promoter and the operator); together they constitute the *lac* operon. Repressor mutations have been isolated that inactivate the repressor protein, and operator mutations are known that prevent repressor-operator binding; such mutations cause continuous production of *lac* mRNA and are said to be constitutive.

When lactose is cleaved by β-galactosidase, the products are glucose and galactose. When glucose is present in a growth medium, the *lac* enzymes are unnecessary. The following general mechanism for preventing transcription of many sugar-degrading operons has evolved in bacteria. High concentrations of glucose suppress the synthesis of a small molecule, cyclic AMP (cAMP). Initiation of transcription of many sugar-utilization operons requires binding of a particular protein molecule, called CRP, to a specific region of the promoter for the operon. Binding occurs only after CRP has first bound cAMP and formed a cAMP-CRP complex. Only when glucose is absent is the concentration of cAMP sufficient to produce cAMP-CRP and hence to permit transcription of the sugar operons. Thus, in contrast with a repressor, which must be removed before transcription can begin (negative regulation), cAMP is a positive regulator of transcription.

The operon responsible for degradation of galactose (the *gal* operon) has two adjacent promoters for the same genes. This operon is negatively regulated by a repressor and positively regulated by cAMP-CRP, as with the *lac* operon. However, only one of these promoters, the stronger one, is regulated by cAMP-CRP. Galactose serves two functions in *E. coli*. One function is as a carbon source, when galactose is present and glucose is absent. In this role, the strong promoter should be subject to repression and to regulation by cAMP-CRP, for the same reasons as with the *lac* operon. However, one of the *gal* enzymes is also required for synthesis of a component needed to form the cell wall. If synthesis of this enzyme were inhibited by glucose, cell wall formation would stop; thus, this enzyme must be made constitutively. This need is satisfied by providing a strong promoter that can be shut down and a weak promoter that is always available.

In the system for degrading the sugar arabinose, a positive regulator is activated when arabinose is present, and this regulator turns on transcription. The *ara* operon is also regulated by the cAMP-CRP system.

Biosynthetic enzymatic systems exemplify another type of regulation. Such a system operates most efficiently when the concentration of each component is determined by the amount of the reaction product in the growth medium. For example, in the synthesis of tryptophan, transcription of the genes encoding the *trp* enzymes is controlled by the concentration of tryptophan in the growth medium. The resulting rate of enzyme synthesis is always sufficient to maintain a rate of tryptophan synthesis necessary for adequate overall protein synthesis. Amino acid-synthesizing operons are usually regulated by attenuator systems—transcription is initiated continually but terminated at a site ahead of the genes encoding the enzymes. The frequency of termination of transcription is determined by the availability of tryptophan; with decreasing concentration of tryptophan, termination occurs less often and the tryptophan-synthesizing enzymes are made, thereby increasing the concentration of tryptophan.

Regulons are complex systems in which synthesis of more than one mRNA molecule is regulated. The HTP regulon consists of a set of proteins that are made when cells are stressed by exposure to high temperature. These proteins are involved in macromolecular synthesis. The Pho regulon is responsible for uptake and utilization of inorganic phosphate when phosphate is limiting. It consists of at least 24 genes, which form several operons. The SOS regulon mediates a complex response to DNA damage. The RecA protein becomes a protease when DNA is damaged. This proteolytic activity cleaves the LexA protein, a repressor both of the *recA*, *uvr*, and SOS genes. In this way, DNA damage activates the major systems for repair of DNA damage.

Production of ribosomes is one of the major synthetic processes in bacteria. As cells grow and divide, the number of ribosomes per cell is balanced with the availability of a carbon source. Transcription of the three rRNA genes and synthesis of the 52 different ribosomal proteins are regulated.These synthetic processes are regulated by two feedback systems. In the first, ribosomes are made in slight excess over that which is needed for a particular growth rate and the free nontranslating ribosomes inhibit the synthesis of rRNA. This is called ribosome feedback regulation. In the second system particular r-proteins inhibit translation of the various r-protein mRNA molecules. Thus, if all ribosomes are in use, more rRNA will be made; the free r-proteins will be utilized in forming new ribosomes and more r-proteins will be made. If a cell is transferred to a growth medium in which growth is slower, many ribosomes will not be used, rRNA will not be made, r-proteins will be in excess, and no more r-proteins will be made.

The activity of enzymes is also frequently regulated, in order to avoid wasted consumption of energy and precursor molecules. In an unbranched biochemical pathway, commonly the product of the reaction sequence inhibits the first enzyme in the pathway. This inhibition is called end-product or feedback inhibition. In a branched pathway the end product of each branch usually inhibits the first enzyme following the branch. Occasionally, in branched pathways isoenzymes (enzymes that catalyze the same reaction but respond to different inhibitors) catalyze the first step in the pathway. When this occurs, the end product of each branch usually inhibits one of the isoenzymes, so that intermediates in the pathway are synthesized in quantities appropriate for synthesis of the products that are not present.

In a few cases, changes in gene expression occur that seem to be random rather than as a response to the concentration of some molecule. An example is the site-specific inversion that causes phase variation in *Salmonella*. In this system, a DNA segment containing the promoter for both the *H2* flagellin gene and an inhibitor of the *H1* gene can assume two orientations. In one, H2 flagellin is made and the inhibitor prevents transcription of the *H1* gene. In the other orientation neither *H2* is transcribed nor is the *H1* repressor made, so the *H1* gene is transcribed. Inversion does not occur in response to any known environmental signal.

There is no universal principle of regulation of gene expression, either in prokaryotes or eukaryotes. Evolution has probably merely selected a mechanism that works for each situation.

BOLD TERMS

adenylate cyclase, aporepressor, attenuator, background constitutive synthesis, basal synthesis, β-galactosidase, cAMP, cAMP-CRP, catabolite-sensitive operons, constitutive, coordinate regulation, CRP, cyclic AMP, cyclic AMP receptor protein, derepression, end-product inhibition, feedback inhibition, flagellin, galactokinase, galactose epimerase, galactose transferase, gene expression, gratuitous inducers, heat-shock proteins, HTP regulon, inducer, inducible, isoenzymes, isozymes, leader, negative regulation, operator, operon, permease, phase variation, positive regulation, regulon, relaxed, repressor, site-specific inversion, stringent factor, stringent response, thiogalactoside transacetylase

ADDITIONAL INFORMATION

Experience in teaching has indicated that one of the most difficult things for a student learning about regulation is the determination of the expression of an operon in a partial diploid when there are cis-dominant mutations. This is an important skill, since analysis of partial diploids is a standard way to distinguish positive and negative regulation and to distinguish structural genes from sites in the DNA. Thus, as a learning aid, a few examples will be worked out for the *lac* operon.

Example I: $lacZ^+lacY^-/lacZ^-lacY^+$. In this simple partial diploid there are no cis-dominant mutations. Note that the location of the two genotypes (that is, to the left or the right of the slash) is unimportant, since expression from a chromosome or a plasmid is identical. We ask whether β-galactosidase is made and, if so, whether its synthesis is constitutive or inducible. The genotype at the left is capable of making the enzyme, whereas that on the right is not; nonetheless, the cell will contain enzyme, as long as the operon can be turned on. To turn it on, functional regulatory elements need to be present, as they are, and an inducer must enter the cell. The latter can be done because the genotype at the right can be induced to make the permease. Thus, for this partial diploid β-galactosidase can be made, and its synthesis is inducible.

Example II: $lacO^{c}lacZ^{-}lacY^{+}/lacZ^{+}lacY^{-}$. This partial diploid has the cis-dominant mutation $lacO^{c}$. Note that the genotype at the left contains the constitutive mutation $lacO^{c}$, so the genes at the left are expressed. However, from the *lacZ* gene at the left a defective enzyme is made. At the right is a functional *lacZ* gene from which active enzyme can be made. However, its synthesis is under the control of a normal operator (note that the operator genotype is not indicated, so by convention it is $lacO^{+}$) and hence enzyme synthesis must be induced. Thus, the partial diploid makes a defective enzyme constitutively and a normal enzyme by induction, so altogether the cell can make the enzyme by induction.

Example III: $lacP^{-}lacZ^{+}/lacO^{c}lacZ^{-}$. At the left is a promoter mutation, which is cis-dominant. This means that no *lac* mRNA can be made from this segment of DNA, and the genes may be thought of as being absent. That is, the genotype of the cell can be considered to be just that at the right, which makes a defective enzyme constitutively, as in Example II. Thus, there is no way for the cell to make active enzyme.

Example IV: $lacI^{+}lacP^{-}lacZ^{+}/lacI^{-}lacZ^{+}$. The genotype at the left contains a promoter mutation, so the *lacZ* and *lacY* genes can be considered absent. However, the *lacI* gene has its own promoter, so *lac* repressor molecules will be present in the cell. The genotype at the right could alone make enzyme constitutively because of the *lacI* mutations. However, a functional *lacI* product is made by the genotype at the left. Thus, the cell can make enzyme, and synthesis must be induced.

Example V: $lacI^{+}lacP^{-}lacY^{+}/lacI^{-}lacY^{-}$. This genotype differs from that in Example IV by the presence of a $lacY^{-}$ mutation on the right. Again, the genotype at the right contributes only *lacI* repressor to the cell, so any synthesis of the enzyme must be inducible. However, the genotype at the right is $lacY^{-}$, indicating that inducer cannot enter the cell. Thus, this cell is unable to make any enzyme, because synthesis would have to be induced and there is no way to induce it.

The synthesis of enzymes by addition of substrate molecules or other nutrients is excellent evidence that the enzymatic system is inducible. A simple and now-classic experiment indicates whether the system is negatively regulated. If a *lac* operon were placed in a cell lacking repressor, transcription would occur. This situation can be created by transferring the *lac* operon from an Hfr male cell (Chapter 2) having the genotype $i^{+}z^{+}$ to a female cell that is $i^{-}z^{-}$ (which makes neither repressor nor β-galactosidase). When the *lac* DNA enters the female cell, which contains no repressor, *lac* mRNA would immediately be made and β-galactosidase would appear in the female. Since i^{+} mRNA would also be synthesized in the female from the transferred *lac* DNA, repressor would also be

synthesized in the female (though at a much lower rate than for β-galactosidase). Thus, the synthesis of enzyme would not continue indefinitely. This phenomenon has been observed for all negatively regulated operons.

Recently a novel type of translational regulation has been observed in bacteria: regulation by antisense RNA. This is an RNA molecule that is complementary either totally or in part to an mRNA. It can hybridize with mRNA and thereby prevent translation. In *E. coli* antisense RNA has been observed to be transcribed either from the strand complementary to that which produced the mRNA or from a duplicated copy of the sequence.

Please note that in the following, for simplicity, the notation *i, o, z, y, and a* may be used for the more cumbersome symbols *lacI, lacO, lacZ, lacY, and lacA.*

DRILL QUESTIONS

1 Is the regulation of gene activity in prokaryotes mainly transcriptional or post-transcriptional?

2 What is meant by coordinate regulation?

3 Other than the ability to turn on and off a set of genes in an operon by a single regulatory element, what else is accomplished by having a set of genes contained in one polycistronic mRNA molecule?

4 Which type of regulation, positive or negative, involves removal of an inhibitor?

5 Would synthesis of an enzyme that is needed continually be regulated?

6 What is the biochemical action of an inducer?

7 Is a repressor necessarily inactivated by an inducer?

8 Is an inducer necessarily an inactivating agent?

9 What is meant by a partial diploid?

10 Physically what is the consequence of binding of the *lac* repressor to the *lac* operator?

11 Which enzymes of the *lac* operon are regulated by the repressor?

12 What term describes a gene that is expressed continually, even though its trans-

cription may be autoregulated?

13 Is the partial diploid $F'i^+/i^-$ inducible or constitutive?

14 Is the partial diploid $F'o^+/o^c$ inducible or constitutive?

15 Why are all constitutively synthesized proteins not made at the same rate?

16 Is the *lac* repressor itself made constitutively or is it induced?

17 Is it necessary for a repressor gene to be adjacent to the operator?

18 Is it necessary for the operator to be very near the promoter?

19 When glucose is present, is the concentration of cyclic AMP high or low?

20 Can a mutant with either an inactive adenylate cyclase gene or an inactive *crp* gene synthesize β-galactosidase?

21 Does the binding of cAMP-CRP to DNA affect the binding of a repressor in any way?

22 Are all proteins translated from a single polycistronic mRNA necessarily made in the same quantity?

23 A gene product is made constitutively. What two features of the system control the rate of its synthesis?

24 Repressors and aporepressors both bind molecules that are components of the metabolic pathway encoded in an operon. What is different about the positions in the pathway of the molecules bound, and what is different about the activity of the complex formed when binding occurs?

25 Is the attenuator, like the operator, a binding site?

26 Is RNA synthesis ever initiated at an attenuator?

27 Antitermination and attenuation are both concerned with termination of transcription. How do they differ with respect to the role of RNA polymerase?

28 Are all ribosomal proteins translated from a single polycistronic mRNA?

29 What molecules regulate translation of r-protein mRNAs?

30 If a cell culture is transferred from a "slow" growth medium to a "fast" growth medium, will the number of ribosomes per cell increase or decrease?

31 In feedback inhibition of a single unbranched pathway, which substance is the inhibitor and which enzyme is inhibited?

32 In a feedback-inhibited pathway with two branches, if isoenzymes were utilized in regulation the activity of the pathway, how many isoenzymes would normally be used?

ANSWERS TO DRILL QUESTIONS

1 Transcriptional and exerted at initiation and termination (attenuation) steps.

2 Turning on and off a set of genes as a unit, simultaneously or by a single signal.

3 The ratio of the amount of each component is maintained constant without having each one separately regulated.

4 Negative regulation.

5 Not always, especially if needed in large quantities. If it were regulated, it would usually be autoregulated.

6 It binds to a repressor.

7 Yes, by the definition of a repressor.

8 No. In a negatively regulated system it is an inactivator. In a positively regulated system it is an activator, but not of a repressor.

9 A cell in which two copies of a small number of genes are present.

10 The promoter is inaccessible to RNA polymerase.

11 β-Galactosidase (*lacZ* product), lactose permease (*lacY* product), and the transacetylase (*lacA* product).

12 It is constitutive.

13 Inducible, because the *lacI* product is diffusible.

14 Constitutive.

15 The major reason is that all promoters do not have the same strength, so the rate

of initiation of transcription varies from one gene to the next.

16 Constitutively.

17 No, because the repressor is a diffusible protein.

18 Yes, since binding of the repressor to the operator interferes with binding of RNA polymerase to the promoter.

19 Low.

20 No. In the first case, it could not make cAMP. In the second, it could not make CRP.

21 No

22 No. There is often a polar effect in that the relative amounts of each protein decrease toward the 3' end of the mRNA.

23 The strengths of the promoter and the ribosome binding site. The latter may be determined by nearby sequences.

24 With a repressor the molecule is usually an early (generally the first) reactant in the pathway. The repressor is inactivated by the binding. With an aporepressor the molecule is usually the product of the pathway. An aporepressor is activated by the binding.

25 No. Nothing is specifically bound to an attenuator.

26 It is strictly a potential termination site for transcription.

27 In attenuation the structure of the termination sequence is altered so that RNA polymerase either does or does not see it. In antitermination RNA polymerase is altered so that it can ignore certain termination sites.

28 No.

29 A single r-protein encoded in a particular r-protein mRNA regulates translation of that mRNA.

30 Increase, because the rate of ribosome synthesis will increase.

31 The product of the pathway is usually the inhibitor, and the first enzyme in the pathway is usually inhibited.

32 Two.

PROBLEMS

1 The glycolytic pathway is responsible for the degradation of glucose, one of the most fundamental energy-producing systems in living cells. Would you expect the enzymes of this pathway to be regulated? If so, in what way?

2 Is it fair to say that a constitutive gene is unregulated?

3 Is it necessary in a bacterial operon for the gene for a repressor to be near the structural genes?

4 Explain how lactose molecules first enter an uninduced $i^+z^+y^+$ cell to induce the *lac* operon.

5 For each of the following diploid genotypes, indicate first whether β-galactosidase can be made, and second, whether synthesis of the enzyme is inducible (I) or constitutive (C) and finally, whether or not each cell could grow with lactose as sole carbon source.

(a) $i^+z^-y^+/i^-z^+y^+$.
(b) $i^+z^+y^+/o^ci^+z^-y^+$.
(c) $i^+z^-y^+/o^cz^+y^+$.
(d) $i^+z^+y^-/i^-z^-y^+$.
(e) $i^-z^+y^-/i^-z^+y^+$.
(f) $i^-z^+y^+/i^+o^cz^-y^+$.
(g) $i^+p^-z^+/i^-z^-$.
(h) $i^+o^cz^-y^+/i^+z^+y^-$.
(i) $i^+p^-o^cz^-y^+/i^+z^+y^-$.
(j) $i^-p^-o^cz^+y^+/i^-z^+y^-$.

6 A cell that is wild-type with respect to the *lac* operon (+ for all alleles) is in a growth medium containing neither glucose nor lactose; that is, it is using another carbon source. How many proteins are bound to the DNA comprising the *lac* operon? How many if glucose is present?

7 Describe in terms of mRNA synthesis, enzyme synthesis, and enzyme activity what happens when lactose is added to a Lac$^+$ *E. coli* culture previously grown in a medium lacking all sugars. Assume that the amount of lactose added is consumed after two generations of growth

8 Repeat the preceding problem for a Lac$^+$ culture in lactose-containing medium when an amount of glucose is added that is consumed in one generation.

9 An $i^+o^+z^+y^+$ Hfr culture is mated with an $i^-o^+z^-y^-F^-$ culture. In the absence of any inducer in the medium β-galactosidase is made for a short time after the Hfr and F^- cells have been mixed. Explain why it is made and why only for a short time. What would happen if the female were $i^+o^cz^-y^-$?

10 You have isolated a Lac$^-$ mutant and found by genetic analysis that its genotype

is z^+y^+. The mutation, which you call i^* is in the *lacI* gene. The partial diploid $i^*z^+y^+/i^-z^+y^+$ is constructed and its phenotype is found to be Lac$^-$—that is, i^* is dominant. The diploid $i^*z^+y^+/o^ci^+z^+y^+$ is Lac$^+$ (β-galactosidase is made). Suggest a property of the mutant repressor that would yield this phenotype. Would $i^*z^+y^+/o^ci^+z^-y^+$ make the enzyme?

11 The purpose of regulating gene expression is to reduce wasted synthesis, because cells experiencing waste are at a definite disadvantage in the long run. Approximately 5% of the total protein made by i^- cells, whose synthesis of β-galactosidase is unregulated, is β-galactosidase, far more than is needed. (a) Suppose 10^6 wild-type and 10^6 i^- cells are placed in a growth medium lacking lactose. If the cells are grown for 50 generations (with continued dilution of the culture to keep the cells growing), what would you expect to happen to the ratio of wild-type and i^- cells? (b) A gratuitous inducer is an inducer that is not a substrate of β-galactosidase. Many experiments are carried out with these inducers to avoid effects of changing concentration of lactose. Suppose a Lac$^+$ culture is grown for hundreds of generations in the presence of a gratuitous inducer; what type of spontaneous mutant might accumulate in the culture?

12 A mutant strain of *E. coli* is found that produces both β-galactosidase and permease whether lactose is present or not. (a) What are two possible genotypes for this mutant? (b) Another mutant is isolated that produces no β-galactosidase at any time but produces permease if lactose is present in the medium. If a partial diploid is formed from these two mutants, in the absence of lactose neither β-galactosidase nor permease is made. When lactose is added, the partial diploid makes both enzymes. What are the genotypes of the two mutants?

13 Will *lac* mRNA be induced by lactose if the *lacZ* and *lacY* genes are deleted and the *lacI* gene and operator are intact?

14 How many proteins are bound to the *gal* operon when (a) galactose is present and glucose is not, (b) both galactose and glucose are present, and (c) neither sugar is present.

15 What is the state of the *gal* operon following addition of lactose to a Lac$^+$ cell? (Remember the chemical structure of lactose and the reaction catalyzed by β-galactosidase).

16 How many proteins are bound to the *trp* operon when (1) tryptophan and glucose are present, (2) tryptophan and glucose are absent, or (3) tryptophan is present and glucose is absent?

17 An *E. coli* mutant is isolated that is simultaneously unable to utilize a large number of sugars as sources of carbon. However, genetic analysis shows that each of the operons responsible for metabolism of each sugar is free of mutation. What genotypes of this mutant are possible?

18 An operon has the gene sequence *A B C D E*. Neither the promoter nor the operator have been located. The repressor gene maps very far away from the structural genes. Various deletion mutants have been isolated. Some deletions of gene *E*, but none for any of the other genes, result in constitutive production of the mRNA of the operon. Where do you think the operator and the promoter are?

19 If an operon is derepressed and then, at a later time, repressed again, will the ability to carry out the enzymatic reactions determined by the structural genes of the operon be eliminated as rapidly as transcription is turned off?

20 An operon responsible for utilizing a sugar Q is regulated by a gene called *kyu*. When Q is added to the growth medium, Qase is made; otherwise, the enzyme is not made. If the gene *kyu* is deleted (denoted Δkyu), no Qase can be made. The partial diploid $kyu^+/\Delta kyu$ is inducible. Two types of point mutants of *kyu* are found: *kyu1*, which never makes Qase, and *kyu2*, which makes the enzyme constitutively. The partial diploids $kyu^+/kyu1$ and $kyu^+/kyu2$ are inducible and constitutive, respectively. What is the likely mode of action of the protein encoded by the *kyu* gene?

21 An operon involved in utilizing a sugar X is regulated by a gene called *b*. When X is added to the cells, Xase is made; otherwise it is not. If the *b* gene is deleted (this is denoted by Δb), no Xase can be made. The diploid $b^+/\Delta b$ is inducible. Point mutants of *b* are of two types; *b1* never makes Xase, *b2* is constitutive. The partial diploids $b^+/b1$ and $b^+/b2$ are inducible and constitutive, respectively. What is a likely mode of action of the protein encoded by the *b* gene?

22 A particular operon seems to be regulated by a gene called *p*. It has been suggested that the *p*-gene product is not a protein but an RNA molecule that is not translated. The principal support for this idea is that a protein product has never been detected. A series of *p* mutants is isolated and characterized. Some of these are temperature-sensitive mutants. In an effort to study these thoroughly, the mutations are transferred into strains carrying various suppressors. In two strains, one carrying an amber and one an ochre suppressor, a particular p^- mutation behaves as if it is p^+. You now know whether the *p*-gene product is a protein or RNA molecule. Which is it and how do you know?

23 In a hypothetical *E. coli* operon the regulator gene is closely linked to a region containing two structural genes (consider only one) and an operator. In the table at the end of the problem (see next page) the regulator, the operator, and a structural gene are listed in correct sequence. The ability of each of the indicated genotypes to synthesize an enzyme under induced and noninduced conditions is

as shown. Which of these genes is the regulator, the operator, and the structural gene? Explain the reasons for your selection.

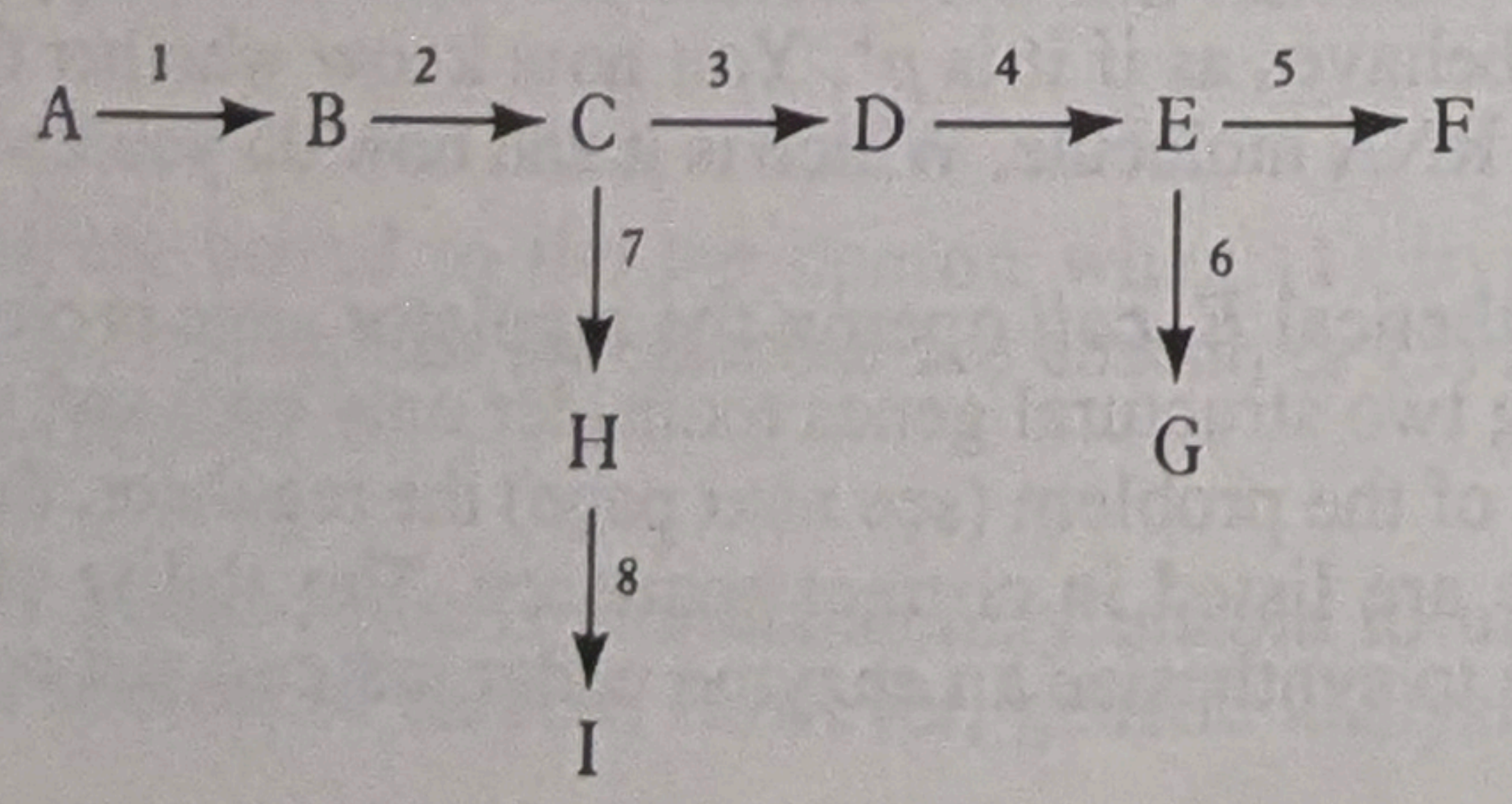
Genotype	Phenotype	
	Inducer absent	*Inducer present*
$a^- b^+ c^+$	S	S
$a^+ b^+ c^-$	S	S
$a^+ b^- c^-$	s	s
$a^+ b^- c^+/a^- b^+ c^-$	S	S
$a^+ b^+ c^+/a^- b^- c^-$	s	S
$a^+ b^+ c^-/a^- b^- c^+$	s	S
$a^- b^+ c^+/a^+ b^- c^-$	S	S

Note: S = enzyme synthesized in normal quantities; s = little or no synthesis.

24 The regulation of an operon responsible for synthesis of X is dependent on a repressor, a promoter, and an operator. In the presence of X, the system is turned off; an interaction between X and the repressor forms a complex which can bind to the operator. (a) What kinds of mutations might occur in the repressor? Describe their phenotype (in terms of whether the operon is on or off). (b) Describe the phenotype of a partial diploid with one wild-type and one mutant gene for each mutant gene.

25 Explain the fact that if a cell culture is transferred from a "fast" growth medium to a "slow" growth medium, the number of ribosomes in the cells that exist at the time of transfer remains the same. What change will occur in subsequent generations?

26 Consider the branched pathway regulated by feedback inhibition, shown below. (a) Which enzymes are subject to feedback inhibition? For each identify the inhibitor. (b) Which steps are likely to be catalyzed by a set of isoenzymes and how many isoenzymes are probably in these step?



ANSWERS TO PROBLEMS

1 Since it is needed continuously, it need not be repressed by substrates or positive effectors. However, the amount of all proteins should be regulated in order that there is neither runaway nor inadequate synthesis. Systems of this sort are often autoregulated in that the rate of synthesis of the enzymes depends on the amount of enzyme present. Thus, as a cell grows and enlarges, the concentration of the enzymes drops and more enzyme is made in order to maintain a constant concentration.

2 No. Constitutivity simply means that transcription of the gene is turned on all of the time. The rate of transcription may be modulated by autoregulation, processing of the RNA may be regulated, and there may be some mode of translational control.

3 No. The repressor is a diffusible protein, so the gene can be anywhere. In fact, there are many repressors, of which the *trp* repressor is one example, whose genes are located quite far from the structural genes. The only reason that a repressor gene is frequently near the structural genes is that in the course of evolution they will tend to approach one another to reduce the probability of being separated by recombination, translocation, and inversion.

4 At very low levels all inducible operons are transcribed—for example, a repressor might come off the operator for an instant and RNA polymerase will get on. Hence each cell will contain a few β-galactosidase and permease molecules. The permease will bring in a few lactose molecules and these will be converted to the true inducer, allo-lactose; then derepression can occur.

5 (a) Yes, I, yes; (b) yes, I, yes; (c) yes, C, yes; (d) yes, I , yes; (e) yes, C, yes; (f) yes, I, yes; (g) no, neither, no; (h) yes, I, yes; (i) no, neither, no; (j) yes, constitutive, no.

6 If neither glucose nor lactose is present, two proteins (the repressor and CRP-cAMP) are bound to the DNA. If glucose alone is present, only one protein (the repressor) is bound to the DNA.

7 First, *lac* mRNA and, then, β-galactosidase is made. When the lactose is finally consumed, synthesis of *lac* mRNA will cease, and no more enzyme will be made. The enzyme already made will persist, so the enzymatic activity remains; however, the enzyme concentration drops as it is diluted out by cell division.

8 Synthesis of *lac* mRNA stops shortly after glucose is added. Enzyme is made for several minutes until the *lac* mRNA is degraded. Enzyme activity is unchanged because the enzyme is stable. After one generation the glucose is consumed and synthesis of *lac* mRNA begins.

9 The *lac* genes are transferred by the Hfr cell and enter the recipient. No repressor is present in the recipient, so *lac* mRNA is made. However, a *lacI* gene also is transferred to the recipient, so soon afterward repressor will be made and *lac* transcription will stop. If the female has the genotype $i^+o^cz^-y^-$, repressor is present and no enzyme is made. The o^c mutation in the female is irrelevant.

10 The cell cannot be derepressed. The simplest explanation is that the repressor cannot bind the inducer. The partial diploid given at the end of the problem would not make enzyme, because the constitutive component makes no enzyme and the other component cannot be induced.

11 (a) In general, when a substance X is lacking from the medium, an enzyme that degrades it has little value to the cell. Cells in which 20% of their protein is useless must grow slightly more slowly. If the culture medium does not limit growth for other reasons, growth will be limited by how fast the necessary quantities of the required proteins can be made. The repressorless constitutive strain should grow 20% slower; that 20% disadvantage will be compounded in each generation. When the inducible strain has undergone 30 doublings, the constitutive one will have undergone only $(0.8)(30) = 24$ doublings. At that time, for every 2^{30} cells of the inducible strain, there are only 2^{24} cells of the constitutive strains, that is, 2^6 times as many cells of the former as of the latter. Therefore, the ratio of wild-type (inducible) cells to repressorless (constitutive) cells is $2^6 = 64$. (b) A noninducible mutant, owing to a repressor that is defective in binding the gratuitous inducer. The reasoning is the following. The gratuitous inducer is not a substrate of the enzyme; that is, it is not a carbon source. Ability to remain uninduced would confer a growth advantage on a cell in these special conditions, so if such a mutation arose spontaneously, it would be selected for in these circumstances.

12 (a) The strain is constitutive, that is, i^- or o^c. (b) The second mutant must be z^- since no β-galactosidase is made; also, it must have a normal repressor-operator system, since permease synthesis responds to lactose. That is, its genotype is $i^+o^+z^-y^+$. Since the partial diploid is regulated, the operator in the operon with a functional gene must be wild-type. Hence, the first mutant must be i^-.

13 Very poorly, because there can be no permease to get the lactose into the cell and no enzyme to form allo-lactose.

14 (a) Two—cAMP-CRP and RNA polymerase. (b) None. (c) Two—repressor and cAMP-CRP.

15 Lactose is cleaved to glucose and galactose, so the *gal* operon is derepressed. The presence of glucose depresses the cAMP concentration so that transcription is from *S2*. If the amount of lactose were limiting, when the glucose is depleted, transcription would be from *S1* and would occur if there were sufficient galactose to maintain the derepressed state. This would probably be the case since the concentration of the *gal* enzymes is not very high when the *S2* promoter is used. Thus, most of the galactose is metabolized after the *lac* operon is derepressed, when lactose is limiting. If there is ample lactose, the *gal* operon will never be operating at maximal level.

16 Glucose is irrelevant to the functioning of the *trp* operon. When tryptophan is present (1, 3), the repressor is active and is bound to the DNA.

17 It clearly lacks a general system that regulates sugar metabolism. The only thing we know about is the cAMP-dependent system, namely, that several mutations prevent activity of this system. For example, the mutant could make (1) either a defective CRP protein, and either fail to bind to the promoter or be unresponsive to cAMP, or (2) an inactive adenyl cyclase gene, making no cAMP.

18 The operator is probably next to gene *E*. Deletions that remove the operator make transcription constitutive. The promoter cannot be between the operator and gene *E*, because in that case all deletions that remove the operator would also remove the promoter, and no constitutive deletions would be found. The order is *E o p*, which is consistent with the fact that some deletions are not constitutive; these would be larger deletions that include both the operator and the promoter. Thus, the gene order (transcription from left to right) is *p o E D C B A*.

19 No. There will be no further synthesis of the enzymes but the enzymes will persist. In time (but a long time) the enzymes might be degraded. If the cell continues to divide, the enzyme activity will gradually decrease, owing to distribution of the enzyme molecules among daughter cells.

20 Since addition of Q causes Qase to be made, the system is regulated. Deletion of *kyu* prevents induction, suggesting that *kyu* encodes a positive regulatory element. A partial diploid with both wild-type *kyu* and the deletion is regulated, indicating that the *kyu* product is present and that it is a positive regulator activated by an inducer. Examination of mutants confirms this view. The *kyu1* mutant never makes Qase but a partial diploid with wild-type is inducible, indicating that the *kyu1* mutation is recessive. It seems likely that the *kyu* product binds Q and is activated. The *kyu2* mutation is constitutive and dominant, suggesting that it is able to turn on transcription of Qase without Q. The *kyu1* product probably fails to bind Q and the active positive regulator is a Q-Kyu complex. The *kyu2* product is probably able to bind to the promoter without first binding Q.

21 It is a positive regulator activated by an inducer. The *b1* mutant cannot be activated. The *b2* mutant is activated without an inducer.

22 A gene product must be a protein if it is inactivated by a chain termination muta-

tion. Suppression by both an amber and an ochre suppressor indicates that the mutation is a chain termination mutation.

23 Site *a* is an operator, *b* is the structural gene, and *c* is a regulator; *a*$^-$ is an operator-constitutive mutant.

24 (a) Three kinds of repressor mutations might occur: (1) a repressor that cannot bind X—the operon is on; (2) a repressor that cannot bind to operator even when X is present—the operon is on; (3) a repressor that binds to the operator without X—the operon is off. (b) Phenotypes of partial diploid with wild-type and each of the mutants are: first mutant, inducible; second mutant, inducible; third mutant, unable to synthesize X under all conditions.

25 The number will stay the same because ribosomes are not destroyed. The concentration will decrease because growth will continue at a rate that exceeds the rate of synthesis of ribosomes. When cell division occurs, the number of ribosomes per cell will be less than there would have been if the medium shift had not occurred, so the average number of ribosomes per newborn cell will decrease. As these cells grow, the total number of ribosomes per cell will not reach the number in fast-growing cells.

26 (a) G inhibits 5; H inhibits 7; G and H together or E alone probably inhibit 3; J inhibits 8; enzyme 1 could be inhibited by (G, H, and J), (E and J), or (C and E); ( ) enclose the products that act together. (b) Step 1, three isoenzymes; step 3, two isoenzymes.

MORE CHALLENGING PROBLEMS

1 What type of mutation of the *lac* repressor might be both constitutive and trans-dominant?

2 Although β-galactosidase and transacetylase are coordinately regulated, their polypeptide chains are not produced in equimolar amounts. The half-lives of the relevant portions of the mRNAs are: β-galactosidase, 90 sec; transacetylase, 5 sec. The average interval between translational starts on one mRNA for the two proteins are: β-galactosidase, 3 sec; transacetylase, 16 sec. Calculate the relative rates of production of the two polypeptide chains.

3 In *E. coli* the enzyme alkaline phosphatase (APase) is coded for by the gene *phoA*. If phosphate, a product of the reaction catalyzed by APase, is added to a cell culture, synthesis of APase is greatly inhibited in the growing cells. There are two other genes, *phoB* and *phoR*, which also seem to be involved in regulating the expression of *phoA;* mutations in *phoB* lead to reduced synthesis of APase in the presence or absence of phosphate. Mutations in *phoR* lead to constitutive synthesis of APase (see table below). We know that the *phoB* and *phoR* genes each produce a protein, so the mutations shown in the table do not represent cis-acting mutations. We also know that the *phoA* gene is being regulated by transcriptional control. Assume that no other proteins or small molecules besides the ones mentioned here are involved in regulating the *phoA* gene.(a) Which one of the regulatory proteins is interacting directly with phosphate to regulate the expression of APase? (b) Which one of the regulatory proteins is interacting directly with a site on the DNA? Is this protein a positive or negative effector? (c) Explain how a complex between the PhoB and the PhoR proteins might function in regulating the expression of APase. Include the role of phosphate in your explanation.

	Relative level of APase when phosphate is	
Mutants	*Absent*	*Present*
phoA$^+$ *phoB*$^+$ *phoR*$^+$	100	20
phoA$^-$ *phoB*$^+$ *phoR*$^+$	0	0
phoA$^+$ *phoB*$^-$ *phoR*$^+$	5	1
phoA$^+$ *phoB*$^+$ *phoR*$^-$	100	100
phoA$^+$ *phoB*$^-$ *phoR*$^-$	5	5

Note: Consider the *phoB*$^-$ to be a point missense mutation in the *phoB* gene.

4 In one bacterial species the histidine operon is a negatively regulated system containing ten structural genes that encode enzymes needed to synthesize histidine. The repressor protein is also coded within the operon—that is, in the polycistronic mRNA molecule that encodes the structural genes. Synthesis of this mRNA is controlled by a single operator regulating the activity of a single promoter. The corepressor of this operon is tRNA$^{\text{His}}$ to which histidine is attached. This tRNA is not encoded in the operon itself. A collection of mutants having the properties indicated below have been isolated. For each of the mutants (*a–e*) determine (i) whether the histidine enzymes would be synthesized by each of these mutants. and (ii) whether each mutant would be (1) dominant, (2) cis-dominant only, or (3) recessive to its wild-type allele in a partial diploid. (a) The promoter cannot bind RNA polymerase. (b) The operator cannot bind the repressor protein. (c) The repressor protein cannot bind to DNA. (d) The repressor protein cannot bind histidyl-tRNA$^{\text{His}}$. (e) The tRNA$^{\text{His}}$ by itself (that is, without histidine) can bind to the repressor protein.

ANSWERS TO MORE CHALLENGING PROBLEMS

1 Since the repressor is a multisubunit protein, a mutant repressor that could not bind to the operator and could bind tightly to a good repressor subunit, thereby inactivating it, would be constitutive and trans-dominant.

2 The two proteins are made from the same polycistronic mRNA. The rate of production of each polypeptide chain is proportional to the product of its mRNA lifetime and the rate of initiation of protein synthesis. Thus, the ratio of the rates of production of a β-galactosidase chain and a transacetylase chain is (90 sec x 1/3 sec^{-1}) divided by (55 sec) (1/16 sec^{-1}) or 8.8.

3 (a) The PhoR protein, because in its absence the expression is insensitive to the presence or absence of phosphate. (b) The PhoB protein; it is a positive effector. The PhoB protein must be binding to DNA and acting as a positive effector necessary for synthesis of APase. If the PhoR protein were acting directly on the DNA, then it would have to be a repressor. Mutants of the *phoR* gene do show constitutivity, as repressor mutants do, but a *phoB*$^-$*phoR*$^-$ double mutant would also show constitutivity if the *phoR* gene were a direct repressor, and they do not. Therefore, since the PhoR protein cannot be a direct repressor of transcription, the PhoB protein must be a direct positive control element affecting transcription of the *phoA* gene. (c) If the PhoB protein binds to the DNA at a promoter-like spot, and the PhoR protein in the presence of phosphate acts to reduce the synthesis of the PhoA protein, then the PhoR protein probably forms a complex with the PhoB protein and phosphate, making it harder for the PhoB protein to bind at the promoter site. Thus, transcription is inhibited. The *phoB*$^-$ mutation allows only 5% expression of the *phoA* gene in the absence of phosphate. In the presence of a *phoR*$^+$ allele the residual expression is still sensitive to phosphate expression (the expression is reduced to 1%) but in the presence of the *phoR*$^-$ allele the 5% expression of APase cannot be reduced by phosphate.

4 (a) (i) No. (ii) 2. If the promoter cannot bind RNA polymerase, the adjacent gene cannot be transcribed. (b) (i) Yes. (ii) 2. If the operator cannot bind the repressor, there will be no repression and transcription of adjacent genes cannot be shut off. (c) (i) Yes. (ii) 3. The repressor cannot bind to the operator, so the system is always on; however, in a partial diploid in which a nondefective repressor is made, this repressor can bind to the operator and prevent transcription. (d) (i) No. (ii) 1. The repressor cannot be inactivated. It will bind to all *his* operators and prevent all his transcription. (e) (i) No. (ii) 1. The tRNA cannot be prevented from activating the repressor. Therefore, all *his* operators will be occupied and all *his* transcription will be prevented.

CHAPTER 16

REGULATION IN EUKARYOTES

CHAPTER SUMMARY

Regulation of genetic systems in eukaryotes is accomplished in a variety of ways, most of which are quite different from the mechanisms observed in eukaryotes. In higher organisms the environment of the cells is fairly constant, and cells are instead usually called upon to respond to signals coming from within the organism. The response may be a temporary one, such as synthesis of a digestive enzyme, or a permanent change to a differentiated cell. Eukaryotic genes are rarely arranged in operons consisting of adjacent genes, but form more loosely organized systems called gene families. Genes in a family are usually not adjacent and may even be located on different chromosomes. Several genes in a family can often be activated simultaneously; it is assumed that each gene has a common regulatory region that responds to a particular signal, though signals of this type have not yet been identified. Some gene families consist of multiple copies of a single gene. In developmentally regulated gene families the individual genes are expressed at different stages of differentiation.

Differentiation and synthesis of particular eukaryotic gene products (either protein or RNA) are induced in several different ways, the method depending on the requirements of the cells. For example, the enormous amount of ribosomal RNA needed during development of an amphibian egg is obtained by specific replication of the rRNA genes, thereby producing a huge number of these genes, each of which can be transcribed; this phenomenon is called gene amplification. Two modes of gene amplification have been observed. In the one used for amphibian rRNA, the gene is excised and circularized, the circles multiply by θ replication, and each circle ultimately enters the rolling circle replication mode. For the chorion gene system gene amplification occurs by *in situ* replication in the chromosome, in which a replication origin is repeatedly activated. The production of huge amounts of silk fibroin during cocoon formation by moths is accomplished not by gene amplification but by synthesis of a long-lived mRNA molecule, so that translation can continue for several days.

Many genes are regulated by turning transcription on and off as needed, as is the case for prokaryotes. For example, many hormones stimulate transcription. The mechanism is poorly understood but invariably involves a hormone-receptor complex on the cell membrane. Presumably, either a macromolecular complex containing the hormone binds to a site on the chromosome and acts as a positive regulator, or the hormone complex removes a negative regulator; while these ideas seem reasonable, direct experimental confirmation is not yet available for many systems. In yeast transcription is regulated by the binding of specific proteins to DNA sites called upstream activation sites. The regulation can be of the positive type, as in the galactose-induced activation of the GAL4 product, which binds to the upstream site of the *GAL1* gene. The yeast *ADR* genes seem to be controlled by a negatively regulated system. Genetic evidence supports the idea of a repressor and an operator, but no repressor has been identified. In plants some transcription is controlled by light via the mediator phytochrome. A light-sensitive transcription factor may be involved.

Base sequences upstream from some eukaryotic promoters are required for initiation of transcription. Examples are upstream activation sites of yeast, upstream sites, and enhancers. Transcription factors are needed for the initiation of transcription of most, if not all, eukaryotic genes. Upstream sites may be binding sites for transcription factors. The expression of some genes is mediated by activation or deactivation of specific transcription; however, no mechanisms have yet been clearly elucidated. In some cases, regulation may be indirect in that programmed synthesis of a particular transcription factor causes a particular gene to be transcribed at a specific time, that is, when the transcription factor reaches a critical concentration.

A controversial subject is the role of DNA methylation in gene expression. In the higher organisms genes that are inactive are often more highly methylated than active genes. Furthermore, some inactive genes can be activated by growth of cells in media containing 5-azacytidine, an inhibitor of methylation. On the basis of 5-azacytidine data it has been suggested that one copy of the human X chromosome is inactivated by methylation. The controversy is based upon lack of direct chemical measurement of the degree of methylation, the fact that some methylated genes are active, and that methylation is totally absent in some oganisms. The idea of methylation as a mechanism for permanent turning off of a gene is an attractive one because a simple mechanism exists for inheriting a methylation pattern. Methylation occurs in a C in an adjacent CG dinucleotide (embedded in certain surrounding sequences), but only occurs if the CG in the complementary strand is already methylated. Thus, once methylation of a parental DNA molecule occurs, each daughter strand will be methylated in precisely the same positions as those in the parent molecule.

Differential processing of primary RNA transcripts is an important mechanism for regulating synthesis of particular proteins. For example, in different cell types unique patterns of splicing may produce two mRNA molecules having distinct leader sequences for a particular gene, resulting in different efficiencies of translation in each cell type. Different patterns of splicing also have been observed to produce distinct proteins from the same primary transcript. Regulation of the site of addition of poly(A) termini can also yield different mRNA molecules from a single primary transcript.

In prokaryotes the relative concentrations of functionally related proteins are usually maintained by the encoding of each protein in a single polycistronic mRNA molecule. Different efficiencies of translation of each gene determine the rate of synthesis of each protein. However, eukaryotic mRNA is not polycistronic, as only a single protein can be translated from a eukaryotic mRNA molecule. In eukaryotes, equal concentrations of related proteins are frequently obtained by synthesis of a long polypeptide called a polyprotein. The amino acid sequences of several proteins may be contained in the polyprotein; hence, cleavage of the polyprotein at particular sites yields distinct proteins. The individual proteins are usually present in equimolar quantities, since one copy of each amino acid sequence is typically contained in the polyprotein.

Translational control, which is rare in prokaryotes, is fairly common in eukaryotic cells. For instance, preexisting mRNA molecules in sea urchin eggs cannot be translated until the RNA is activated by the recruitment factor. In organisms in which reticulocytes are anucleate and hence lack DNA, synthesis of globin, the principal protein product of these cells, is balanced with that of hemin in that the concentration of hemin regulates overall protein synthesis.

Certain cells of the immune system are programmed to produce particular immunoglobulins. The specific immunoglobulin that is made is determined by a complex sequence of DNA splicing events. There are about 300 different *V* genes, which encode the first 95 amino acids of the variable region of the IgG L chain, and four different *J* genes. In each precursor a single recombinational event occurs in which all copies of the *V* gene, but one, and all copies of the *J* genes, but one, are eliminated. The result is a segment of DNA that contains one *V* gene, one *J* gene, and the *C* gene, which encodes the constant region of the IgG L chain. A similar sequence of events produces a single coding sequence for the H chain of IgG. This recombination occurs in an embryonic cell, which goes on to multiply and produce a clone of cells capable of making only a single L chain and a single H chain.The different features of the immune system that enables an organism to synthesize nearly 10^8 different antibodies may be summarized as follows. (1) Germline diversity: each haploid genome contains nearly 500 *V*, *J*, and *D* segments corresponding to the H and L chains; (2) combinatorial joining: nearly 800 different *V–J* units for L chains and 8000 different *V–J–D* units can be formed by splicing; (3) junctional variation: shifts in the junction points of the *V–J* and *V–D–J* junctions produce additional diversity; (4) somatic mutation: in a completely unknown way mutations occur in the *V* regions and these increase the possible number of antibody molecules; (5) random association of H and L chains: any pair of light chains can join with any pair of H chains to produce the two antigen-binding sites, which are jointly formed at the ends of the Y-shaped molecule by the V_L and V_H domains.

BOLD TERMS

antibody, antigen, cell cap, clonal selection, constant region, endocytosis, estrogen, 40-kd transcription factor, gene amplification, gene family, gene loss, gene rearrangement, H chain, hemin-controlled repressor, hormones, housekeeping genes, IgG, immunoglobulin, immunoglobulin G, L chain, masked mRNA, macronucleus, micronucleus, moderately prevalent RNA, myelomas, phytochrome, polyprotein, polytene, pseudogenes, recruitment factor, superprevalent mRNA, totipotent, translational amplification, V region, variable region, vitellogenin

DRILL QUESTIONS

1 Are the individual members of a gene family necessarily coordinately regulated?

2 Can the members of most gene families be cloned in a single plasmid?

3 Are all members of a gene family on a single chromosome transcribed in the same direction?

4 Are all members of a gene family actively transcribed in all tissues of a single organism?

5 Why are related eukaryotic genes rarely regulated by controlling synthesis of a single polycistronic mRNA, as is common in prokaryotes?

6 What biochemical process is usually regulated by hormones that control synthesis of a particular gene product?

7 Do hormones ever bind directly to DNA?

8 Name two mechanisms utilized in eukaryotes for maintaining a fixed ratio of gene products, when equality of the products is not required.

9 With reference to Drill Question 8, what mechanism are utilized when equality of the gene products is required? Give two examples.

10 What regulatory strategies are used when only a limited time is available for synthesis of a large amount of a protein?

11 Distinguish gene amplification from translation amplification.

12 How does gene amplification of the rRNA genes and of the chorion genes differ?

13 What regulatory strategies are used when a long time is available for production of a particular protein?

14 In a eukaryote an extracellular regulator forms an intracellular complex that ultimately binds to DNA. What barriers must it pass in order to reach the DNA?

15 In prokaryotes regulatory elements are in fixed positions with respect to the gene(s) regulated. How does the situation differ in eukaryotes?

16 What is meant by housekeeping genes?

17 If an enhancer were moved from a position near gene *A*, where it has strong enhancing activity, to a position 50 nucleotides upstream from gene *B*, which is constitutively transcribed, would transcription of gene *B* be increased?

18 In what ways can differential processing regulate synthesis of particular gene products?

19 How many chains and of how many different types are present in an IgG molecule?

20 What is meant by the variable and constant region of the individual chains of an IgG molecule?

21 Distinguish germ line cell and antibody-producing cell.

22 Do the *V* and *J* segments of an antibody gene join to form the variable or the constant region of an IgG molecule?

23 How many genes are linked to form the H chain of an IgG molecule?

ANSWERS TO DRILL QUESTIONS

1 No. A gene family may consist of related genes that are expressed fairly independently of one another, for example, the developmentally regulated families.

2 No. This is fairly rare since the member genes are frequently located throughout a chromosome or on different chromosomes.

3 No.

4 No.

5 Eukaryotic genes are not generally organized into polycistronic operons, because without complex splicing events, only one protein molecule can be translated from a primary transcript.

6 Transcription.

7 No examples are known and it is quite unlikely. The active form is probably a hormone-receptor complex, which may bind directly to DNA or may activate another binding protein.

8 (1) Differential processing of a single primary transcript. (2) Common regulation of the synthesis of different transcripts.

9 The general mechanism is to break down a precursor into subunits, each of which is one of the desired products. One example is the cleavage of an RNA precursor to obtain rRNA components. A second example is cleavage of a polyprotein.

10 Gene amplification is used when a limited time is available, because by increasing the number of genes, the rate of production of a particular protein per cell can be increased greatly.

11 In gene amplification the number of copies of a gene is increased; therefore, more mRNA can be made per unit time. In translational amplification, the lifetime of the mRNA is increased, which enables a great deal of mRNA to be present at a particular time.

12 Ribosomal RNA genes: the gene is excised from the DNA and it replicates first as a circle by the θ mode and then as a rolling circle. Chorion genes; replication occurs *in situ* by repeated activation of a replication origin.

13 When a long time is available, achieving a high rate of synthesis is not necessary; instead, continued synthesis of mRNA and extended lifetime of mRNA accomplish the end.

14 First, it must pass through the external cell membrane. Then, it must pass through the nuclear membrane. Finally, it may have to disrupt nucleosome structure and remove histones, before it can actually reach the DNA.

15 Of course, regulatory elements are in fixed positions, except for certain transposable elements described in Chapter 21. However, the difference is that in eukaryotes the regulatory elements can be experimentally moved, often by hundreds of nucleotide pairs, without affecting their regulatory activity.

16 Genes that provide for general cell function and which are transcribed and translated continuously.

17 Very likely, unless gene *B* was somehow separated from some element needed for its activity.

18 (1) Different promoters can be spliced to different genes. (2) Different AUG codons can be made proximal to the 5′ terminus of the mRNA. (3) Different stop codons can be placed in phase with a start codon. (4) Different coding sequences can be produced by removal of particular introns. (5) Different sites of polyadenylation can determine whether a particular nearby segment can be excised.

19 There are four chains, two H chains and two L chains.

20 All IgG molecules do not have the same amino acid sequence. The constant region has nearly the same amino acid sequence in all IgG molecules, whereas the amino acid sequence of the variable region differs in each type of IgG molecule.

21 The germ line cell contains the nucleotide sequence found in sperm and egg. A germ line cell does not make antibody. The antibody-producing cells, each of which makes a specific antibody, are the progeny of a germ line cell in which excision and splicing of segments of the IgG gene components has occurred.

22 Variable region. A *V* segment and a *J* segment are joined to the *C* region of the gene. The *C* region encodes the constant region.

23 Four. These are the *V*, *J*, and *D* segments, which are joined to the constant region.

PROBLEMS

1 What methods are available in eukaryotes to turn on synthesis of a set of proteins at the same time?

2 What information about gene organization is obtained by observing that several members of a gene family can be simultaneously cloned in a single plasmid?

3 In discussing gene families one frequently mentions embryonic, fetal, and adult forms of a protein. To what phenomenon do these terms refer? What term is used to describe a gene family of this sort?

4 Name three mechanisms that increase the number of gene copies per cell.

5 Name several possible mechanisms for increasing the concentration of a particular RNA molecule to a very high value (that is, the production of superprevalent RNA).

6 In what way is the regulation of a gene that is active in a differentiated cell inherently different from regulation of a bacterial gene?

7 Why cannot gene amplification produce a permanent change in a cell that can

persist from generation to generation?

8 (a) What role does genetic recombination play in rDNA amplification in *Xenopus* oocytes? (b) Several types of nucleases must be required for rDNA amplification. What are these types? (c) The excess rDNA in *Xenopus* oocytes is not present in cells of an embryo. Explain what has happened to this DNA. (d) The signals that cause the phenomenon in (c) are not known. Suggest a few possibilities.

9 Differentiated cells are often called upon to produce an enormous amount of a particular protein. What methods are used when only a limited time is available for this synthesis and when a long time is available?

10 How might a cell be signaled to synthesize a particular protein but no other?

11 In prokaryotes positive regulatory elements often provide a binding site for RNA polymerase, which strengthens the binding of the polymerase to the transcription start site (e.g., cAMP-CRP and the λ cII protein). What is the major evidence that suggests that this does not occur in eukaryotes.

12 In the yeast system for metabolizing galactose is the *GAL4* regulator a positive or negative regulatory element? Is regulation like that in *E. coli*?

13 Describe a method used in eukaryotes by which distinct amounts of a particular protein could be synthesized, as opposed to a simple on-off system. (An example was given for yeast.) Describe a similar mechanism in prokaryotes (but not attenuation, because attenuation gives a continuous range of activity).

14 In *E. coli* most regulatory elements can be transferred from one gene to another and still retain regulatory activity. For example, any gene can be relocated next to the *lac* operator and promoter, and transcription of the gene will be regulated by the concentration of lactose. Can similar effects be observed in eukaryotes?

15 Two proteins are synthesized initially at the same time because the two RNAs encoding each protein are made in response to the same signal. At a later time, when the transcription signal is no longer present, one of the proteins is still made at nearly the same rate and the other is not made at all. Suggest a means for this temporal programming.

16 What is meant by totipotency?

17 Differentiated cells rarely return to the undifferentiated state. Suggest several mechanisms for maintaining a differentiated state.

18 A particular hormone typically acts on only one or a small number of cell types. What is the usual property of a cell that determines whether a cell does or does not respond to a hormone?

19 A very abundant mRNA is isolated from a eukaryotic cell. You are interested in knowing how many of copies of the corresponding gene there are per cell. You do this by denaturing and renaturing the total cellular DNA, obtaining a Cot curve. At various times of renaturation DNA samples are taken and you determine by DNA-RNA hybridization whether the gene has renatured yet. In this way, the position of the gene on the Cot curve is determined. From the abundance of the mRNA can you predict where on the Cot curve this DNA will be? Explain.

20 (a) Under what circumstances might the concentration of a particular protein be regulated by turning total protein synthesis on and off, as in the case of globin synthesis. (b) Why can globin synthesis not be transcriptionally regulated in reticulocytes?

21 Describe in outline the mechanism for generating antibody diversity.

22 What, in outline, is the mechanism that enables an organism to make a large amount of antibody to a particular antigen and to retain this programming indefinitely?

23 (a) If you wanted to redesign the mouse genome to increase antibody diversity, but by adding the least amount of DNA, which component would you increase? (b) If an organism has 150 different *V* genes, 12 *J* genes, and 3 possible *V–J* joints for L chains, and can make 5000 different H chains, how many different antibodies can be made?

24 When chromatin from somatic cells of *Xenopus* is isolated and transcribed in an *in vitro* protein-synthesizing system, only somatic genes are transcribed. This is true whether the *in vitro* transcription extracts are derived from oocytes or from somatic cells. When the chromatin is washed with 0.6 M NaCl before transcription, this specificity is lost and oocyte genes are also transcribed. Interpret these results in terms of the molecular basis of the specificity.

25 A particular cell makes substance X in response to a hormone H. Prior to exposure of a culture of the cells to H, a large RNA molecule can be found in the nucleus. It is capped and has a poly(A) tail. After addition of H this RNA is not found; a second RNA, which in an *in vitro* protein-synthesizing system makes X, is found in cytoplasmic polysomes. A segment of DNA has been isolated and both RNA molecules can hybridize to this molecule. What mechanism might be used for hormonal control of synthesis of X?

26 Several eukaryotes are known in which a single effector molecule regulates the synthesis of different proteins encoded in distinct mRNA molecules 1 and 2.

Give a brief possible molecular explanation for each of the following observations made when the effector is absent. (a) Neither nuclear nor cytoplasmic RNA can be found that hybridizes to either of the genes encoding molecules 1 and 2. (b) Nuclear but not cytoplasmic RNA can be found that hybridizes to the genes encoding molecules 1 and 2. (c) Both nuclear and cytoplasmic RNA, but not polysome-associated RNA, can be found that hybridizes to the genes encoding molecules 1 and 2.

27 A substance Q is made in a cell culture in response to an external effector E. In the absence of E, a nuclear RNA consisting of about 800 nucleotides can be found that hybridizes with the cloned gene encoding Q. The RNA is capped, has a poly(A) tail and in an *in vitro* protein-synthesizing system directs synthesis of a tripeptide. If E is added to the growth medium, very little of this nuclear RNA is found and an mRNA molecule containing 770 nucleotides is found in the cytoplasm. This mRNA directs synthesis of Q but not the tripeptide. How might E regulate synthesis of Q?

28 A hormone R directs synthesis in cultured cells of a protein S, which contains 174 amino acids. In the absence of R a nuclear RNA with 853 nucleotides can be isolated that is capped and has a poly(A) tail. The RNA hybridizes with the cloned gene encoding S. If the RNA is used in an *in vitro* protein-synthesizing system, a protein is made that contains 240 amino acids, of which the first 23 amino acids have the same sequence as that of S. The remainder of the protein does not resemble S in any recognizable way. If R is added to the culture, the nuclear RNA is not found, but an mRNA with 624 nucleotides, which also hybridizes with the cloned *S* gene, is found. If this mRNA is used in an *in vitro* protein-synthesizing, S is made, but not the protein with 240 amino acids. What might R do?

ANSWERS TO PROBLEMS

1 (1) Each protein may be encoded in a single gene, and each gene may have the same regulatory sequence or the same transcription factor. (2) All proteins could be obtained from a polyprotein. (3) All proteins could be encoded in a single primary transcript that is differentially processed to yield a set of mRNAs, each of which encodes one of the proteins. You can probably think of other mechanisms.

2 First, the genes are clearly on the same chromosome. Second, from the increase in size of the plasmid the maximum separation between the genes is known.

3 There are many instances in which several distinct protein molecules, having like or identical functions, are prevalent at various stages of development of an organism. The differences between the proteins are often only a few amino acids. The different forms of the protein are encoded in distinct genes, each of which is active at a particular stage of development. A collection of such genes comprises a developmentally regulated gene family.

4 Chromatid amplification, which increases the number of copies of all genes; gene amplification by excision of a gene, followed by replication of the excised unit; localized gene amplification by repeated initiation of DNA replication *in situ* .

5 Gene amplification to increase the gene dosage, increase in the lifetime of the RNA, or activation of a strong promoter. An interesting mechanism would be one in which there are two classes of genes producing the RNA; one could exist in small numbers and have a weak constitutive promoter, and the other could exist in large numbers and have a strong promoter whose activity is regulated.

6 Bacterial gene expression is turned on and off, whereas a gene active in a differentiated cell needs to be turned on and probably never has to be turned off.

7 In gene amplification of excised genes the excised units do not contain centromeres and hence cannot segregate in an orderly fashion during subsequent division. In localized gene amplification replication of the remainder of the chromosome would occur before mitosis and the normal gene number would be restored. In chromatid amplification there is no mitosis, because, for some reason, cells with polytene chromosomes do not divide further; they are terminally differentiated.

8 (a) Since the ribosomal genes replicate extrachromosomally, some type of recombinational event is needed to excise the rDNA. (b) Nucleases to catalyze recombinational excision. These might be endo- and exonucleases having additional functions or there might be a site-specific enzyme like the phage λ integrase. Since rolling circle replication is utilized in the amplification, there must be an endonuclease (site-specific) that makes the initiating nick for this mode of replication. (c) The excised rDNA units fail to replicate and are diluted out as cell division proceeds. (d) The signal is probably one that prevents initiation of rolling circle replication. Fertilization might cause synthesis of an inhibitor of the initiating endonuclease referred to in (b), of a repressor protein that makes the replication origin inaccessible, or an exonuclease that attacks the linear part of the rolling circle until the entire structure is degraded.

9 Limited time—gene amplification; long time—increased lifetime of mRNA with or without gene amplification.

10 The gene encoding the protein could have a unique promoter. An effector molecule could cause synthesis of two new proteins, an inhibitor of the cellular RNA polymerase, and a new RNA polymerase that could recognize only that promoter. You may have other ideas.

11 The upstream sites at which regulatory proteins seem to act are too far from the start site for a polymerase to bind both to a regulatory protein and the start site. Furthermore, the upstream sites can be moved somewhat without significantly affecting the rate of transcription.

12 It is a positive element: binding of galactose to the GAL4 protein stimulates transcription. In *E. coli* the *gal* operon is negatively regulated by a repressor.

13 The yeast *CYC1* gene has two upstream activation sites which are separately regulated by the HAP1 and HAP2 proteins. The stimulatory effects of binding to the sites is additive. This is not unlike the *E. coli gal* system, which has two promoters.

14 Yes. Enhancers can act on many, if not most, genes. In yeast the UAS for the *GAL* system (the one that binds the GAL4 product) will put many genes under control of galactose if it is located several hundred bases upstream from a particular gene.

15 Probably the simplest explanation would be that the two mRNAs had different lifetimes. More complicated explanations in which translation of one of the mRNAs is inhibited could also be invoked.

16 An adult cell is considered to be totipotent if it retains the ability to direct embryonic development when transplanted into an embryo. It is thought that such a cell has not lost any embryonic genes; however, since totipotency is experimentally defined by a transplantation experiment, it is possible that some genes present in the embryo, but not necessary for embryonic development, may be lost.

17 A mechanism might exist for amplifying genes yet there may be no mechanism for gene removal; a repressor-like protein, which is synthesized only once in a cell, may be removed and irreversibly activated; a gene rearrangement may occur; a positive regulator may be activated or synthesized yet there may be no way to deactivate it or to prevent its synthesis; there may be a change in chromosome or chromatin structure (chromatin activation).

18 A cell can respond to a particular hormone only if the cell has a binding site for that hormone on the cell membrane.

19 There is no way to predict this because mRNA abundance is not usually related to gene copy number.

20 (a) The cell must not require continuous synthesis of any essential protein. (b) Reticulocytes have no nucleus.

21 An embryonic cell contains a large number of different base sequences that constitute the coding sequences for all antibodies. These sequences are contiguous in the DNA. In the course of development a reciprocal recombinational event joins distant sequences and at the same time removes large blocks of DNA that include many adjacent sequences. There are many different blocks that can be removed, so that different combinations of coding sequences can remain after this recombinational event. One event occurs in a particular cell leaving that cell with a unique coding sequence that enables it to make a particular antibody.

22 First, a cell has engaged in a recombinational event that enables it to make a particular antibody, as described in Problem 21. In a way that is not understood, the cell is also programmed to multiply on continuous (or subsequent) exposure to the antigen. Thus, a clone of cells, which make a single antibody, forms. This response, which is called clonal selection, is an example of cell amplification.

23 (a) Increase the number of *J* genes. (b) $(150)(12)(3)(5000) = 2.7 \times 10^6$.

24 The specificity lies in the chromatin not in the extract and some inhibitor of oocyte transcription is washed from the chromatin by the 0.6 M NaCl.

25 The hormone causes a primary transcript to become smaller and to be able to make a functional gene product. Thus, the hormone is apparently needed to initiate intron excision.

26 (a) The promoters for both transcription units probably have a common sequence acted on by either a positive or negative regulator. (b) Both primary transcripts have a common sequence acted on by an element that prevents some stage of processing. (c) Both processed mRNA molecules have a common sequence involved in ribosome binding. An effector may remove a protein bound to this sequence or it may denature a double-stranded region containing the ribosome binding site.

27 E somehow regulates excision of a single intron that contains an AUG codon that is nearest to the 5' terminus of the RNA. There is also an in-phase stop codon that may or may not be in the intron. Once the intron is excised, there is another AUG codon that initiates synthesis of Q.

28 Directly or indirectly it controls excision of an intron that contains a number of nucleotides that is not a multiple of 3. Excision of the intron changes the reading frame beyond the intron, so a stop codon is encountered that terminates S. If the intron contained 3*n* nucleotides, the reading frame would not change and the C-terminal regions of the two proteins would be identical.

MORE CHALLENGING PROBLEMS

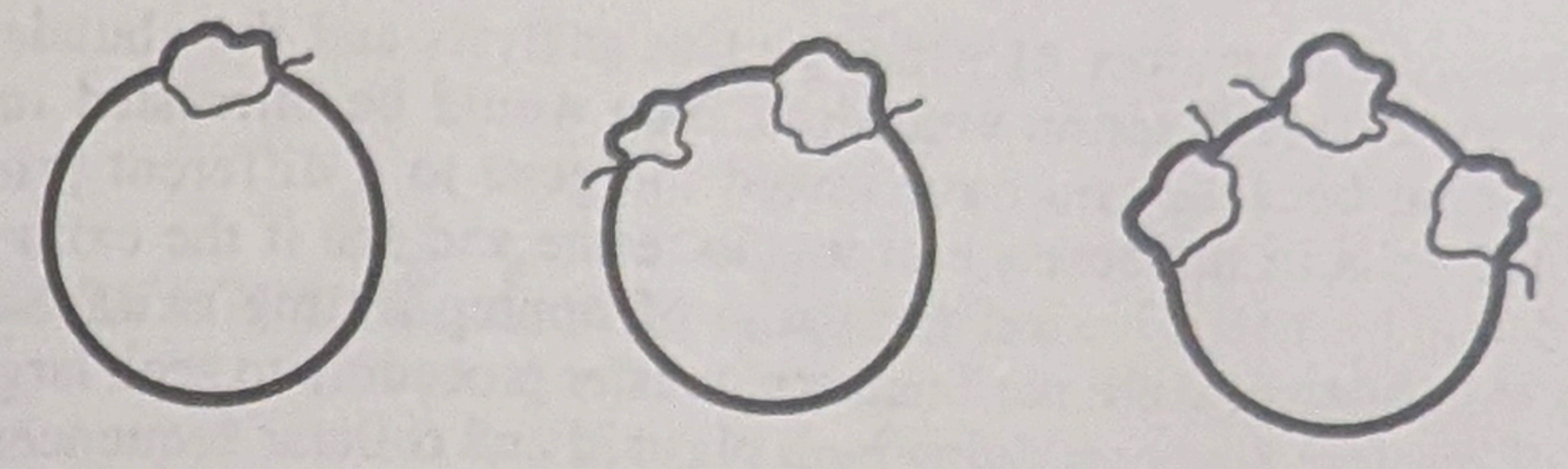
1 A dihydrofolate reductase-deficient (dfr^-) hamster cell line is transformed with

a plasmid containing a *dfr*$^+$ structural gene, linked to the promoter of a hormone-inducible gene, and another gene, *xgp*, which is a dominant selectable marker in this system. Transformed cells were selected for the ability to make Xgp, and all those cells selected also made Dfr. A transformed cell was treated with increasing doses of methotrexate, and a line of cells isolated which is resistant to 200-fold higher levels of methotrexate than the original transformant, and contains 200-fold more Dfr enzyme molecules. (a) How would you determine how much of the increase in Dfr is due to gene amplification and how much to enhanced transcription of the gene? Why would you bother to look at the possibility of enhanced transcription, when you know that previous work with methotrexate-resistant mouse and hamster cells showed that the increased enzyme levels were proportional to (and apparently dependent solely on) the increased gene copy number? (b) If you detected an increased copy number of a gene, how would you determine whether the amplified genes are integrated or extrachromosomal? (c) Would you expect that your methotrexate-resistant cells have increased copies of the *xgp* gene as well as *dfr*, and why or why not? How would you test this possibility?

2 A researcher wishes to study the expression of two kangaroo genes, *A* and *B*, in *E. coli*. In the experiment with gene *A*, she does a partial EcoRI digest of kangaroo DNA, inserts the pieces into the plasmid pBR322, transforms an *E. coli* deletion mutant lacking the *A* gene, and imposes a selection that requires that gene *A* is active. For gene *B* she uses reverse transcriptase to make DNA copies of the mRNA from kangaroo liver cells, which are known to produce large amounts of enzyme B. She then inserts these DNA copies into pBR322, transforms an *E. coli* mutant lacking a functional B enzyme and selects for activity of B. (a) What do you expect the results of these experiments to be and why? (Assume that in both cases the gene sequences have been inserted near a pBR322 promoter and in the proper phase). (b) In a second experiment the kangaroo cells lacking function of the *A* gene were transformed with the plasmid containing gene *A* and kangaroo cells lacking a functional *B* gene were transformed with the plasmid containing gene *B*. The *A* gene but not the *B* gene is expressed. Suggest a possible explanation.

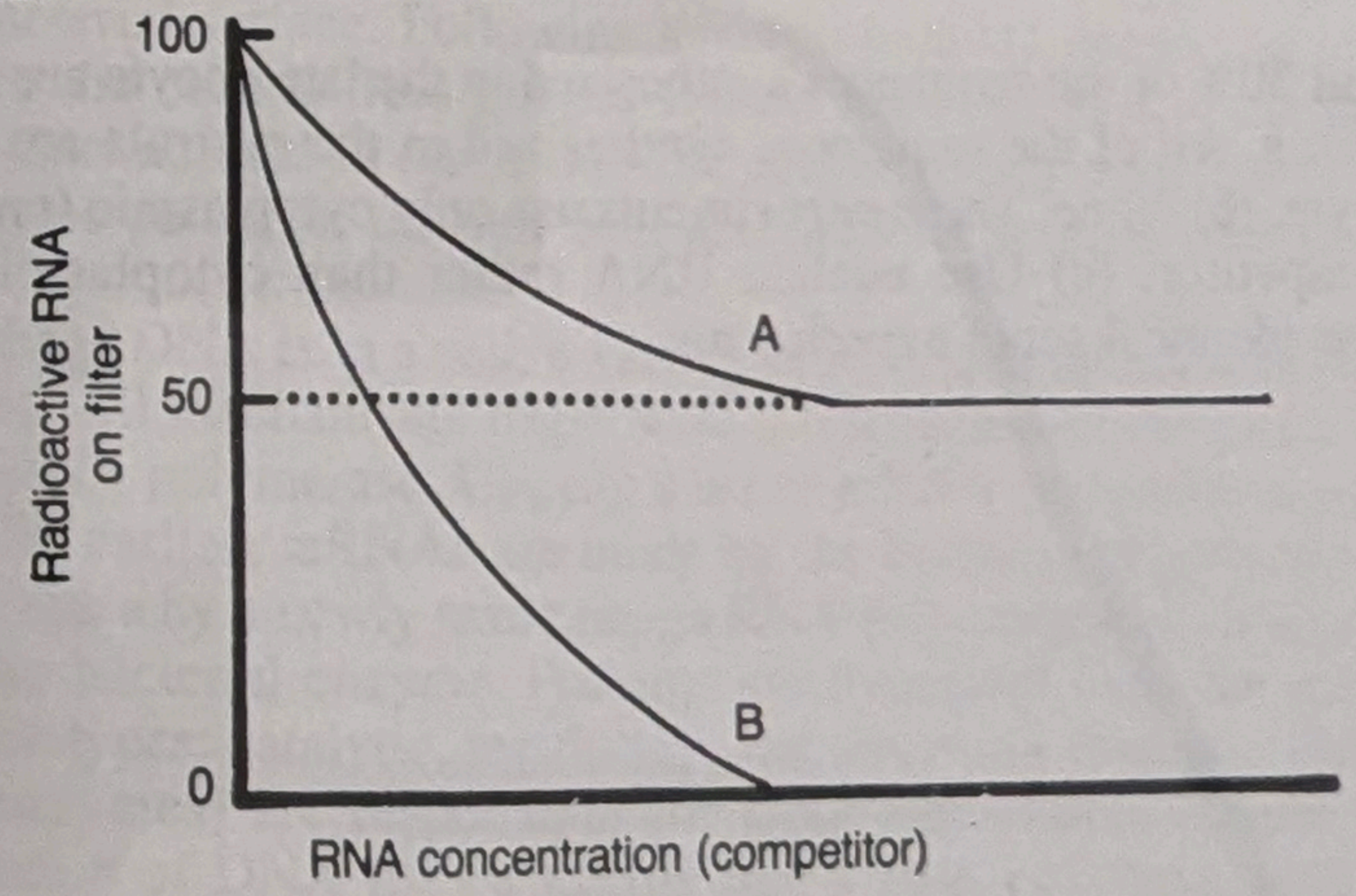
3 A gene family containing three genes is being studied. Messenger RNA molecules for the three genes appear in the cytoplasm at nearly the same time following turn-on of transcription. It has been hypothesized that a single transcript is made and that it is processed in three different ways (perhaps at random to yield the three mRNA molecules). The principal evidence is that the same sequence of four nucleotides is adjacent to the cap in all three mRNA molecules. The entire gene family has been cloned as a single unit in the plasmid pBR322. The plasmid DNA is partly denatured and all three mRNA molecules are added using conditions in which DNA-RNA hybrids form. Thus, some of the plasmid molecules have R loops, that is, bubbles in which one of the DNA strands is joined to mRNA rather than to its complement. The following molecules are observed in electron micrographs of these mRNA-containing plasmids

(other forms are also seen). Could the hypothesis be correct?



4 A particular mRNA molecule is detectable by a cloned c-DNA probe in liver cells but not in brain cells. Chromatin from both cell types is digested with DNase I to a limited extent of digestion (10 to 20% of the DNA rendered acid-soluble) and then hybridized with labeled c-DNA. No sequences are obtained from the liver DNA that hybridize with the c-DNA probe; hybridizable DNA is found after digestion of brain-cell DNA. What can you conclude from these data?

5 *Xenopus* oocytes were synchronously labeled at a late stage of oocyte development with ^{3}H-uridine. Radioactive total cytoplasmic RNA isolated from the oocytes was hybridized with single-stranded filter-bound total DNA from adult *Xenopus*. This RNA:DNA hybrid was competed with unlabeled total cytoplasmic RNA from gastrula-stage *Xenopus* embryos. The result is curve A of the figure below. In a reciprocal experiment, radioactive cytoplasmic RNA obtained from gastrula-stage embryos was hybridized to total DNA on filters. This RNA-DNA hybrid was competed with unlabeled late oocyte cytoplasmic RNA. The result is curve B. (a) What can you conclude about the RNA synthesized at these two stages of early *Xenopus* development? (b) What information does this experiment give you about the role of transcriptional regulation in *Xenopus*? (c) How would you attempt to learn more about the roles of transcriptional regulation in early *Xenopus* development using similar techniques?



ANSWERS TO MORE CHALLENGING PROBLEMS

1 (a) Measure the number of genes by Cot analysis and the abundance of the mRNA by a hybridization procedure. You would be interested in enhanced transcription because you have linked the gene to a different promoter. (b) Grow the cells in the absence of methotrexate and see if the extra copies are diluted out by cell division (one sign of nonreplicating extrachromosomal DNA). Alternatively, use the Southern transfer procedure to seek large pieces of DNA of discrete size containing both plasmid and cellular sequences. (c) Gene amplification usually occurs as amplification of a large segment of DNA so that one would expect both genes to be amplified. Take the *xgp* portion of the plasmid and use as a probe in a Southern transfer experiment. This procedure is useful for amplifying any gene by linking it to the *dfr* gene.

2 (a) The *B* gene may function because it lacks introns. It is unlikely that the *A* gene would function unless it happens to lack introns. (b) It seems as if the gene cannot function without its introns. There is some precedent for this in the globin system: globin mRNA is less stable in the cytoplasm if it is formed by direct transcription of the coding sequence of DNA than if it is spliced from intron-containing RNA.

3 No. Note the single-stranded poly(A) tails projecting from the junctions. These are always at the 3'-OH terminus of the RNA yet they are not at the same ends of each loop. Thus, the RNA molecule of the central gene is transcribed from the strand that is complementary to the coding strand for the outer genes.

4 If a segment of DNA contains a hypersensitive site, it will be digested more rapidly than the bulk of the DNA, so by the time 20% of the DNA is solubilized there will be no DNA from a hypersensitive region remaining. If there is no hypersensitive site, digestion of the DNA of the gene should occur at the same rate as that of bulk DNA. Therefore, the mRNA comes from a region that is hypersensitive in liver cells but not in brain cells.

5 (a) About 50% of the sequences synthesized in the late oocyte are not present in the gastrula. All of the sequences synthesized in the gastrula are present in the late oocyte. (b) None. These experiments use only cytoplasmic (processed) RNA as a competitor. (c) Use nuclear RNA rather than cytoplasmic RNA in an otherwise identical set of experiments.

CHAPTER 17

BACTERIOPHAGES I: LYTIC PHAGES

CHAPTER SUMMARY

A bacteriophage, or phage, is an extracellular nucleoprotein particle that can infect and replicate within a bacterium. A typical phage consists of a protein shell in which is enclosed one double-stranded DNA molecule. A few phages contain single-stranded DNA, or RNA. The most common phage structure is an icosahedral head to which a tail is attached. Less common are tailless heads and filamentous structures.

A phage has one goal: to produce progeny particles. It does so in two ways. In the lytic cycle progeny phage are made during the infection and released from the cell shortly after their assembly. A complete lytic cycle generally occurs in 20–60 minutes at 30–37 C. In the lysogenic cycle a replica of a phage DNA molecule becomes incorporated into the bacterial chromosome. The bacterium survives the infection and the phage DNA replicates as part of the bacterial genome. At a later time, if the bacterium is damaged, a lytic cycle ensues. Both life cycles begin by adsorption of a phage particle to the bacterium; usually adsorption is to the cell wall, but some phages adsorb to appendages (sex pili) from the bacterial surface. Following adsorption tailed phages inject their nucleic acid into the cell; tailless phages release the nucleic acid on the cell surface and the nucleic acid is taken into the cell in poorly understood ways. Filamentous phages penetrate the cell wall and the coat is digested during the process.

Once phage DNA is in a cell, a variety of strategies are used to produce progeny particles. All mechanisms begin with transcription of the phage DNA by the bacterial RNA polymerase. Usually a set of mRNA molecules is made at various times. The earliest mRNAs are made by the bacterial polymerase; later mRNAs may be made by a newly made phage RNA polymerase or various modified forms of the bacterial enzyme. Proteins are translated from the mRNAs. These are of three types: catalytic, regulatory, and structural proteins. The catalytic proteins (enzymes) are responsible for DNA replication, further transcription, preparation of DNA for packaging into a coat, cell lysis, and, occasionally, production of DNA precursors and destruction of bacterial DNA. Regulatory proteins may inactivate bacterial enzymes, but are primarily responsible for causing the various stages of phage production to occur in an orderly and efficient way. Structural proteins form heads, tails, and other components of the finished phage particle. The life cycle usually ends with the bacterium being broken up by a lysing enzyme and release of progeny phage to the surrounding medium. With the M13 class of filamentous phages progeny particles are extruded through the bacterial cell wall without destruction of the cell. The number of phage produced per infected cell (the burst size) depends on the phage species. Typical values are 50–300, but burst sizes of the smallest phages and some of the RNA phages can be in the thousands.

A particular phage type can usually infect only one bacterial species and often only particular strains of that species. The primary limiting factor is the lack of suitable adsorption sites on the bacterium. Additional factors are the inability of a host RNA polymerase to respond to phage promoters and a host restriction system that destroys many kinds of foreign DNA. Host restriction occurs by the activity of a restriction endonuclease that cleaves incoming DNA in particular base sequences (restriction sites). The host DNA may also contain these sites but they are modified by methylation such that they are resistant to the host restriction enzyme. A phage DNA molecule that has replicated in a particular host has been able to do so because its DNA was, in an earlier generation, also methylated in these sensitive sites. A phage that has replicated in a cell that lacks the specific methylating activity will have the restriction site unmethylated and hence its DNA will be destroyed by the host restriction enzyme.

Phages range considerably in size from those that contain fewer than 10 genes and up to 200 genes. The mid-sized and large phages often contain genes that are nonessential in that they are unnecessary in laboratory situations. These genes are believed to be advantageous to the phage in nature. Nonessential genes frequently duplicate host genes and thereby increase the rate at which certain precursor molecules can be synthesized. The phage genes are all organized in several polycistronic mRNAs that are synthesized according to well-regulated time schedules.

E. coli phage T4 is a large phage, whose DNA contains glucosylated 5-hydroxymethylcytosine (HMC), instead of cytosine. The HMC protects the phage DNA against phage-encoded DNases that destroy the bacterial DNA. This destruction provides the phage replication system with a large initial supply of nucleotides. HMC synthesis and glucosylation are carried out by phage enzymes. A phage dCTPase serves to prevent incorporation of cytosine into phage DNA. T4 DNA is terminally redundant and the population of DNA molecules is cyclically permuted. Both features result from the mechanism of DNA replication and packaging. Replication produces long phage-DNA concatemers, which are cut into head-sized units by the packaging system. One headful is greater than one genome by the size of the terminally redundant region. Cutting of the DNA after the redundant segment causes the next DNA molecule to be started at a different point in the genome; this gives rise to the cyclic permutation.

T7 is a midsized *E. coli* phage whose complete nucleotide sequence is known. Of interest is its mode of regulating transcription. The earliest transcript is made by host RNA polymerase. Genes encoded in this mRNA inactivate the host RNA polymerase and synthesize a phage RNA polymerase, which causes the second wave of transcription. The third and final transcription phase is delayed because injection of the DNA into the bacterium is slow, and seven minutes is required before the promoter enters the cell. T7 DNA is linear and replicates according to a mode described in Chapter 9.

λ is a midsized *E. coli* phage whose base sequence is also known. It is the best-understood phage. Its transcription is regulated by several proteins, two of which are antiterminators. The N protein allows termination signals on early mRNAs to be ignored, causing these mRNAs to be extended. From one of the extended mRNAs is made the Q protein; this protein causes antitermination of a short mRNA that becomes the late mRNA, which encodes the structural proteins. DNA replication occurs in two stages. In the first, circles are generated by θ replication. Then, this mode is inhibited and rolling circle replication ensues, producing concatemers that are cut into unit molecules during packaging .

Phage ϕX174 is a small (11 genes) single-stranded circular DNA molecule. It is copied by host enzymes to form a double-stranded circle, called RFI. Looped rolling circle replication of RFI leads to production of progeny single-stranded circles. Transcription occurs from the newly made strand in RFI and is unregulated. The DNA is so small that a large number of replicas are made rapidly (about 2 sec per molecule) and are available by the time the structural proteins have been synthesized in sufficient number for packaging to occur. The lysis enzyme is translated slowly, and lysis does not occur until more than 500 progeny phage have formed.

The filamentous phages, such as M13, are unique in that they do not kill the host. They have only a few genes, which are contained in a single-stranded DNA circle. Replication begins by formation of a double-stranded circle, and rolling circle replication produces progeny. Of special interest is the fact that synthesis of the phage particle occurs within the cell membrane and progeny particles are made as they are being extruded from the cell.

The RNA phages contain three genes that encode an RNA replicase, the coat protein (CP), and an attachment protein (A). The replicase copies the + strand, and the daughter – strand is released. The – strand is then used as a template for synthesis of many + strands. The timing of synthesis and the amounts of the three proteins are translationally regulated. Translation of the *CP* gene opens a double-stranded segment, allowing synthesis of the replicase. As more CP is made, it binds to a site near the replicase gene and blocks further synthesis of replicase. Replication across the *A* gene transiently frees the ribosomal binding site of that gene and allows translation of the gene.

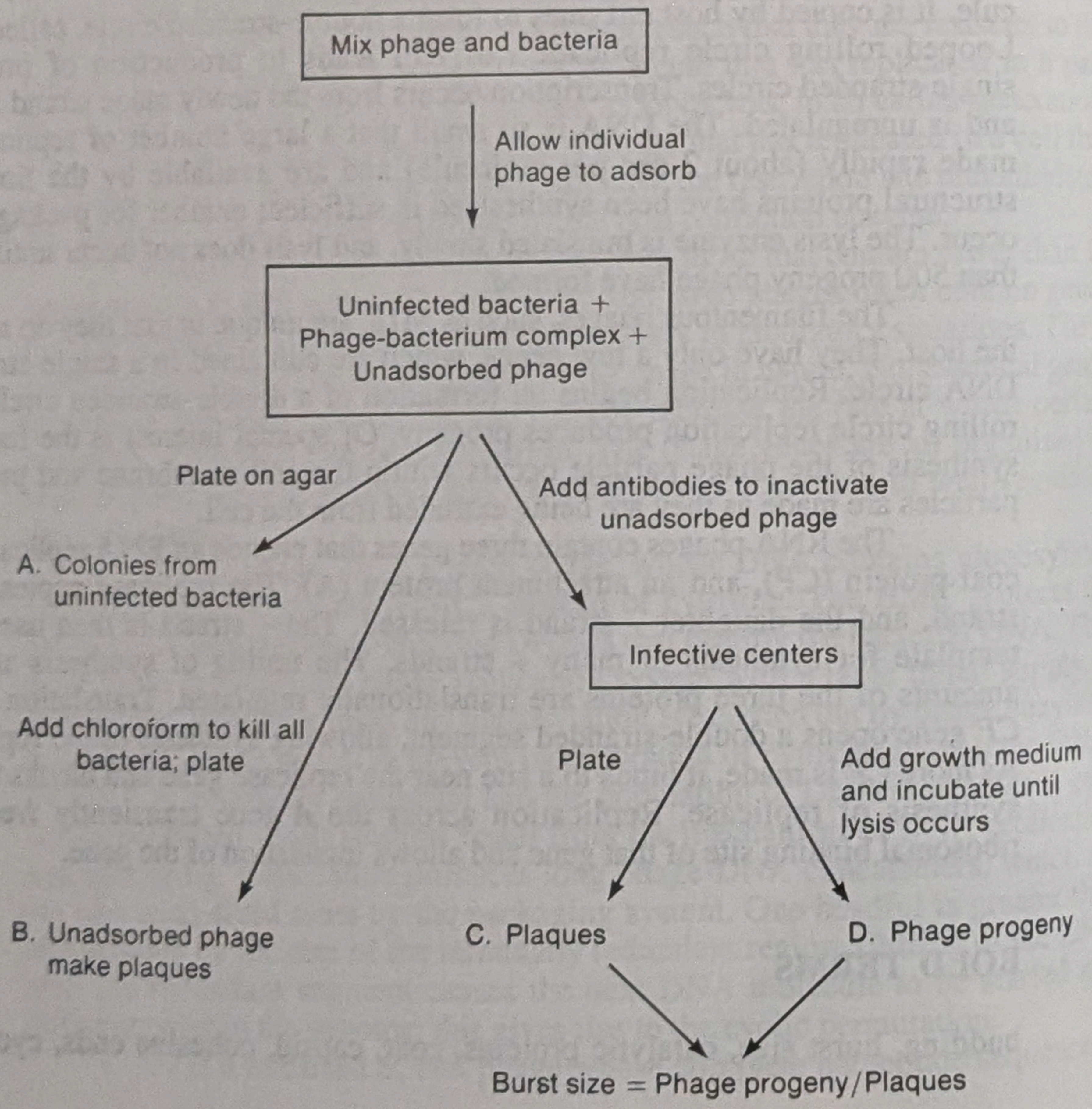
BOLD TERMS

budding, burst size, catalytic proteins, coat, capsid, cohesive ends, cyclically

permuted, early mRNA, early proteins, endolysin, headful mechanism, host modification, host restriction, 5-hydroxymethylcytosine (HMC), infective center, late mRNA, lysis, lysogenic, lysozyme, lytic, maturation proteins, morphogenesis, morphogenetic enzymes, restricted, structural proteins, Ter system, temperate phage, terminally redundant, terminase, virulent

ADDITIONAL INFORMATION

In phage experiments one should know the number of unadsorbed phages and infected cells. In the text, it was shown that the number of infected cells could be determined by adding phage antibody and measuring the number of bacteria that can produce phage. Unadsorbed phage are detected by adding chloroform to the culture after the adsorption period; free phage are unaffected but chloroform kills infective cells, so all plaques formed after chloroform treatment are due to unadsorbed phage. A flow sheet for determining the parameters is given below:


Mix phage and bacteria
Allow individual phage to adsorb
Uninfected bacteria + Phage-bacterium complex + Unadsorbed phage
Plate on agar
Add antibodies to inactivate unadsorbed phage
A. Colonies from uninfected bacteria
Infective centers
Add chloroform to kill all bacteria; plate
Plate
Add growth medium and incubate until lysis occurs
B. Unadsorbed phage make plaques
C. Plaques
D. Phage progeny
Burst size = Phage progeny/Plaques

In a well-designed experiments each of these parameters is measured because often results depend on the MOI.

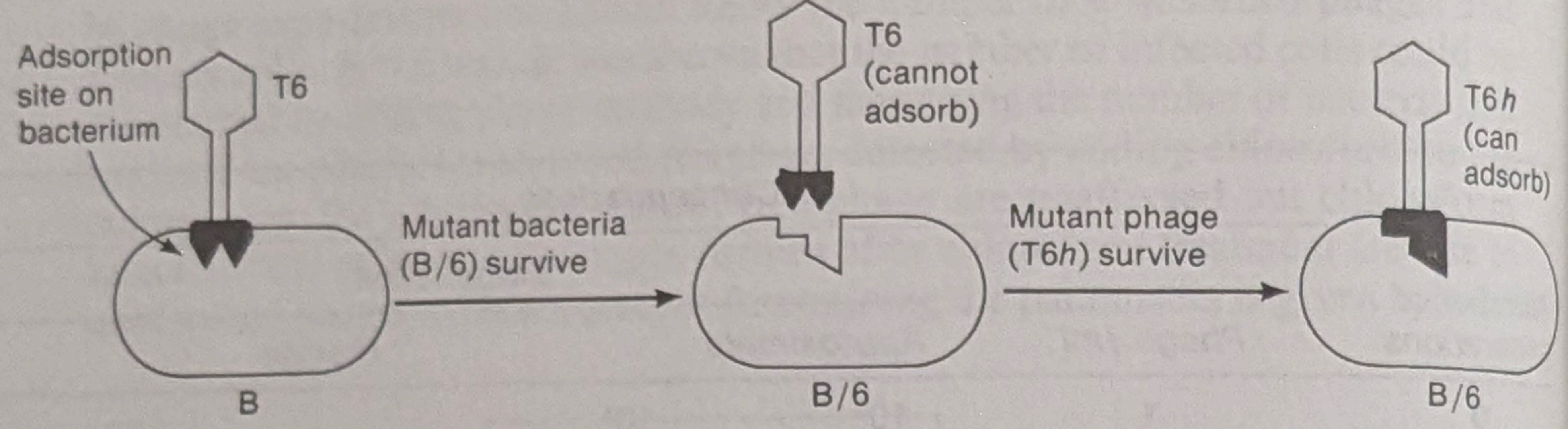
When infecting a bacterial culture, it is often desirable to know when the culture will lyse and what the expected number of phage per unit volume is. This can be simply calculated, as long as the doubling time of the bacteria and the burst size of the phage are known. A sample calculation is shown in the table below in which a single phage whose average burst size is 100 and whose life cycle lasts 25 minutes infects a one-milliliter bacterial culture with a 25-minute doubling time.

Number of generations	Concentrations		
		Bacteria/ml	
	Phage/ml	Approximate	Precise
0	1	10^6	10^6
1	10^2	2×10^6	$2 \times (10^6 - 1)$
2	10^4	4×10^6	$(4 \times 10^6) - (2 \times 10^2) - 4$
3	10^6	7.98×10^6	$(8 \times 10^6) - (2 \times 10^4) - (4 \times 10^2) - 8$
4	10^8	1.4×10^7	$(1.6 \times 10^7) - (2 \times 10^6) - (4 \times 10^4) - (8 \times 10^2) - 16$
5	$1.4 \times 10^9 = 100 \times (1.4 \times 10^7)$	0	0

Initially a bacterial culture at a concentration of 10^6 cells/ml is infected with 1 phage. The doubling time of the bacteria and the life cycle of the phage are equal.

In this calculation, it is assumed that adsorption is always instantaneous and complete. Note that, in four generations, the number of bacteria has increased 14-fold, whereas the number of phage has increased 10^8-fold. At this time there are approximately 8 times as many phage as bacteria; hence, all bacteria are infected. Thus, after 125 minutes the bacteria are gone and the original phage particle has produced 1.4×10^9 progeny.

In the text, the specificity of adsorption was discussed. Some interesting and useful adsorption mutants have been isolated. If 10^8 *E. coli* B cells are infected with 10^{10} T6 phage particles and the infected cells are put on nutrient agar, about 100 colonies will form. Cultures prepared from these colonies will have lost the ability to be infected by T6, and phage will no longer adsorb to the mutant cells. (These cells have of course not been produced by the infection but existed in the culture and were selected by growth with an excess of phage.) These bacterial mutants are called T6-resistant; this is denoted either B/6 (read "B-bar-six") or Tsx-r. Phage mutants can be found that will plate on the resistant cells. For example, if 10^8 Tsx-r cells are used to form a bacterial lawn and 10^8 T6 phage are added, about 10 plaques will result. The phage in these plaques carry a mutation in the tail-fiber gene and have gained the ability to adsorb to Tsx-r cells. These phage are called *h* mutants (for *h*ost range). They usually retain the ability to form plaques on Tsx-s bacteria and hence are said to have an "extended host range." The figure below shows the selection processes schematically. Note that in this figure, for graphic purposes, the tail plate has been modified, although in reality the tail fibers are changed.


Adsorption site on bacterium
T6
B
Mutant bacteria (B/6) survive
T6 (cannot adsorb)
B/6
Mutant phage (T6*h*) survive
T6*h* (can adsorb)
B/6

In the text, it was pointed out that nonessential genes, which seem to be completely unnecessary in the laboratory, have probably survived in the course of time because they are of use in nature. Support for this explanation comes from an observation not described in the text. The *E. coli* λ *b2* segment (which includes several genes) is certainly nonessential in the laboratory, and in fact deletions continually arise, even to the extent of deleting the entire segment. Since λ is not particularly fastidious about the size of the DNA that it packages and since a smaller DNA molecule takes less time to replicate than a larger one, continued cycles of replication of λ usually result in a population in which most of the phages carry *b2* deletions. Since *b2* deletions are not isolated from natural populations, the *b2* segment must contain genes that are essential in nature.

In the text it was emphasized that phages must utilize host RNA polymerase for the initial transcription event since no other RNA polymerase is present at the time of infection. It is essential that a phage set up a program of transcription so that its DNA is transcribed preferentially to the host DNA, and this is usually accomplished by synthesis of a new polymerase (T7) and/or by modification of the host polymerase (T4 and T7). However, *Bacillus subtilis* phage PBS-2 uses a different mechanism. This phage synthesizes a phage-specific RNA polymerase, but the phage RNA polymerase is packaged within the phage head and is injected along with the DNA. This phenomenon is quite rare for phages but, as will be seen in Chapter 23, is not unusual for animal viruses.

DRILL QUESTIONS

1 How many DNA molecules are contained in a typical phage particle?

2 What are the three morphological forms of phage particles?

3 How are phage particles usually counted?

4 How many plaques can be formed by a single phage particle?

5 What is the process by which an infected bacterium releases progeny phage?

6 A phage adsorbs to a bacterium in a liquid growth medium. Before lysis occurs, the infected cell is added to a large number of bacteria, and a lawn is allowed to form on a solid medium. How many plaques will result?

7 A suspension of a phage is exposed for a long time to a DNase. Will the treatment alter the number of plaques that can be produced by the phage suspension?

8 What kinds of nucleic acids are found in phages of different species?

9 What kind of nucleic acid is found in the following phages: T4, T7, λ, ϕX174, M13, MS-2, R17? Tell about circularity also.

10 If 10^6 phage are mixed with 10^6 bacteria and all phage adsorb, what fraction of the bacteria will not have a phage?

11 List the main stages of a phage life cycle.

12 Name the minimal (essential) structural components of the simplest phage particle.

13 What is meant by a nonessential gene? Can a phage gene be truly nonessential?

14 What is meant by the term phage-host specificity, and what is the most frequent cause of this specificity?

15 Why can a particular phage not adsorb to any bacterial species?

16 A particular bacterial mutant cannot utilize lactose as a carbon source. If a phage adsorbs to such a bacterium and the infected cell is put in a growth medium in which lactose is the sole carbon source, can progeny phage be produced?

17 Why do phage plaques not enlarge indefinitely?

18 If a mutation in gene *X* of a particular phage causes a three-fold reduction in plaque size, would gene *X* be considered an essential or a nonessential gene?

19 A sample of a wild-type virulent phage grown on strain A of *E. coli* plates on

strain X with an efficiency of 10^{-4}. What is the most likely explanation for this low efficiency of plating?

20 A sample of a wild-type virulent phage grown on strain A fails to form any plaques (less than 10^{-9}) on strain Y. What is the most likely explanation for the failure?

21 What class of molecules are found in phage tails?

22 Roughly speaking, what is a typical burst size of a phage whose nucleic acid is double-stranded DNA?

23 What is the source of the RNA polymerase that makes the first phage mRNA in the life cycle of a phage containing double-stranded DNA?

24 Considering the enzymes used for transcription, how does transcription of T7 differ from that of λ?

25 What DNA intermediate seems to be obligatory in the replication of single-stranded-DNA phages?

26 Can a phage DNA molecule be terminally redundant without being cyclically permuted?

27 What structural feature is present at the termini of λ DNA that is absent in the DNA of T4 and T7?

28 Following injection, what structural transitions does phage λ undergo, prior to transcription?

29 What two modes of DNA replication are used by λ?

30 What rules, which are different, determine the cutting sites in concatameric T4 and λ DNA?

31 (a) For the single-stranded DNA phages what is meant by the (+) strand and by the (−) strand? (b) Is the meaning the same for the RNA phages? (c) What is the difference with respect to transcription of the (+) strands of the DNA and RNA phages?

ANSWERS TO DRILL QUESTIONS

1 One.

2 Tailed icosahedra, nontailed icosahedra, filaments.

3 In biological experiments one is concerned with the number of viable phage, so they are counted by plating and counting plaques. However, for physical experiments one may need to know the total number of particles, viable and nonviable. This is not a trivial measurement, though a rough idea (no more than a twofold error) can be obtained by electron microscopy. The best method is to measure the DNA concentration of a solution and divide this value by the molecular weight of the phage DNA molecule, if it is known.

4 Just one.

5 Lysis.

6 One. If the bacterium has not lysed, all progeny will be in the same position.

7 No, because DNase cannot penetrate a phage head.

8 Single- and double-stranded DNA (linear or circular); single-stranded RNA and segmented double-stranded RNA (linear only).

9 T4, T7, λ: linear, double-stranded DNA. ϕX174, M13: circular, single-stranded DNA. MS-2, R17: linear, single-stranded RNA.

10 Use the Poisson equation. With a phage:bacteria ratio of 1, the fraction uninfected is e^{-1}.

11 Adsorption, entry of nucleic acid, transcription, production of phage proteins, production of phage nucleic acid, assembly of particles (including encapsidation of nucleic acid), release of progeny particles.

12 A protein coat to enclose the nucleic acid, a component capable of adsorbing to a bacterium, an element designed to enable the nucleic acid to penetrate the cell wall. The latter could be an injection system or a component that stimulates the bacterium to take in the particle.

13 A phage gene is considered nonessential if its product is not needed for growing the phage in the laboratory. Those genes that duplicate host genes may be truly nonessential, even in nature.

14 Phage-host specificity refers to the fact that a particular phage can grow on only one or a small number of bacterial species or strains. The most frequent cause is the inability of a phage to adsorb, except to certain bacteria.

15 In general, the bacterium will lack an adsorption site.

16 No. Phage can only develop in a metabolizing bacterium.

17 A phage can only multiply in a growing bacterium, and the nutrients in the growth medium become depleted.

18 Nonessential, because plaques form.

19 Host restriction.

20 Strain Y probably lacks an adsorption site for λ.

21 Proteins.

22 50 to a few hundred.

23 It must be a host enzyme.

24 After the first transcript is made by the host RNA polymerase, T7 inactivates *E. coli* RNA polymerase and synthesizes a T7-specific polymerase. λ uses the bacterial polymerase throughout the life cycle.

25 A double-stranded circular molecule.

26 Yes. T7 is an example.

27 Complementary single-stranded ends.

28 The linear molecule spontaneously circularizes via the single-stranded termini, and the single-strand breaks are sealed by DNA ligase.

29 θ and rolling circle replication.

30 T4, cut to fill a head; λ, cut at *cos* sites.

31 (a) The (+) is the strand contained in the phage particle or any intracellular strand having the same base sequence. The (–) strand has the complementary base sequence. (b) Yes. (c) The DNA (–) strand is the coding strand, whereas the (+) RNA strand is the coding strand.

PROBLEMS

1 λ phage adsorbs to a site on *E. coli* that is involved in the transport of the sugar maltose. Growth of *E. coli* for many generations in a medium containing maltose enhances the rate of adsorption of λ, and growth in a medium containing glucose completely inhibits adsorption. Suggest an explanation for the role of maltose and the inhibitory effect of glucose.

2 Since infection of a bacterium by a phage is usually lethal to the bacterium, why have bacteria not evolved to lose their phage receptors?

3 Bacteria are allowed to grow on an agar surface until a confluent turbid layer appears. Then, 10^3 T4 phage are spread on the surface. Six hours later (a time sufficient for plaque formation, if the phage had been added at the time the bacteria were placed on the agar) no plaques are evident. Explain.

4 Some types of phages produce small plaques and some produce large plaques. (a) Give several reasons why this might be so, and explain why populations of smaller phages tend to produce larger plaques. (b) Even for a given type of phage there will be varying plaque size if phage particles are added to agar containing bacteria. However, if phage and bacteria are incubated together for several minutes before adding agar, the plaques are more uniform in size. Propose an explanation.

5 If a culture of *E. coli* is infected with T4, the burst size is about 300, whereas with T7 it is about 100. For each, the lysis time is about 22 minutes at 37 C. However, T7 makes plaques whose diameter is roughly four times as great as T4 plaques. Propose an explanation.

6 Phage T4 normally forms small clear plaques on a lawn of *E. coli* strain B. A mutant of *E. coli* called B/4 is unable to adsorb T4 phage particles, so no plaques are formed. T4*h* is a host-range mutant phage capable of adsorbing to *E. coli* B and to B/4 and forms normal looking plaques. If *E. coli* B and the mutant B/4 are mixed in equal proportions and used to generate a lawn, what will be the appearance of plaques made by T4 and T4*h*?

7 In a broth containing glucose and yeast extract, *E. coli* grows with a doubling time of 30 minutes. Phage T7 has a life cycle of 20 minutes under these conditions and a burst size of 200 phage per infected cell. If a culture of 2×10^7 *E. coli* per ml is growing exponentially and 5000 plaque-forming units per ml of T7 phage are added, when will the culture lyse? (In working with this problem assume that phage adsorption is instantaneous and that multiply-infected bacteria give the same burst as singly infected bacteria.)

8 In an experiment in which bacteria and phage are mixed, how does one determine the actual moi and the number of infective centers?

9 One ml of a bacterial culture at 5×10^8 cells/ml is infected with 10^9 phage. After sufficient time for greater than 99% adsorption, phage antiserum is added to

inactivate all unadsorbed phage. The infected cell is mixed with indicator cells in soft agar, and plaques are allowed to form. If 200 cells are put in each petri dish, how many plaques will be found?

10 Soil (earth) has been screened for new phages by washing the soil with buffer and assaying the buffer for plaque-forming units. A fairly ordinary-looking plaque is isolated and the phage in it are used to infect a growing bacterial culture in order to generate a phage lysate that can be studied further. In your preliminary studies you measure the burst size and, remarkably, find that it is 2. A fellow student says that you have obviously made a mistake. What made the student say that? If there is no mistake, what remarkable features must this phage have?

11 P2 and P4 are bacteriophages of *E. coli*. They have the following properties: (1) When one P2 phage infects a bacterium, the bacterium usually bursts, giving about 100 P2 progeny. (2) When a P4 phage infects a bacterium, the bacterium survives because P4 is a defective phage. (3) When P2 phage and P4 phage coinfect the same bacterium, lysis of the bacterium gives 100 P4 progeny and no P2 progeny (because P4 inhibits the growth of P2). If 3 x 10^8 P2 and 2 x 10^8 P4 are added to 10^8 bacteria, then (a) how many bacteria will not be infected, (b) how many bacteria will survive, (c) how many bacteria will produce P2 progeny, and (d) how many bacteria will produce P4 progeny?

12 Give several examples of mechanisms by which a phage converts a bacterium to a phage-producing machine.

13 What would you expect to happen if, after virulent phage infection (such as with T4), all the phage genes were transcribed and translated at once following phage DNA infection?

14 Name four features by which T4 DNA and T7 DNA differ.

15 How do T4 and T7 differ with respect to the DNA and RNA polymerases used in their life cycles?

16 If an *E. coli* culture is simultaneously infected by phages T4 and T7, each at moi = 5, only T4 phage will be produced. From what you know about T4 biology, propose a simple explanation. What would you expect to happen if the T4 was added 10 minutes after addition of T7?

17 What is the function of T4 dCTPase (deoxycytidine triphosphatase) in a T4 infection?

18 Describe the course of T4 phage DNA synthesis following infection of *E. coli* with a T4 mutant which cannot synthesize (a) cytidine hydroxymethylase, or (b) α-glycosyl-transferase.

19 Can one determine whether the early genes or the late genes of phage T4 are injected first into host bacterium? Explain your answer.

20 What is the basic principle used by T4 in regulating the transcription sequence of the mRNA molecules made before DNA replication begins?

21 If a T4 phage has a large deletion which of the following will be true? (1) The activity of one or more proteins will be greatly altered or altogether missing. (2) The phage DNA will be smaller. (3) The terminal redundancy will be larger. (4) The phage DNA will be the same size. (5) Cyclic permutation will be eliminated.

22 Suppose you had a phage whose linear DNA is synthesized by the rolling circle mode and is packaged by "the headful rule" (that is, DNA is added to a head of fixed size until no more DNA can fit). However, the DNA is normally neither terminally redundant nor cyclically permuted. You find a mutant strain of this phage, the DNA of which has a deletion in a nonessential gene. This phage is used to infect a bacterium, and many phage are produced. The DNA is isolated and is treated with an exonuclease, which removes a few bases from the 5'-P end. The treated DNA is then exposed to conditions which could circularize T4 DNA (if it were also pretreated with the exonuclease). This DNA is examined by electron microscopy. Would circles be found?

23 Most double-stranded DNA phage have several classes of mRNA that can be divided into two major groups—early mRNA and late mRNA. The genes carried on the early mRNA species vary from one phage to the next. Nonetheless, there are certain genes that are usually on early transcripts and some which are invariably on late transcripts. What are these genes?

24 Phages T4, T7 and λ use different modes of regulating transcription. What are they?

25 Why is the duration of time allotted to late transcription usually greater than the time for early transcription?

26 What enzymes are needed to form a supercoiled λ DNA molecule after infection? State them in order of activity.

27 Which transcripts of phage λ are made by an N^- mutant?

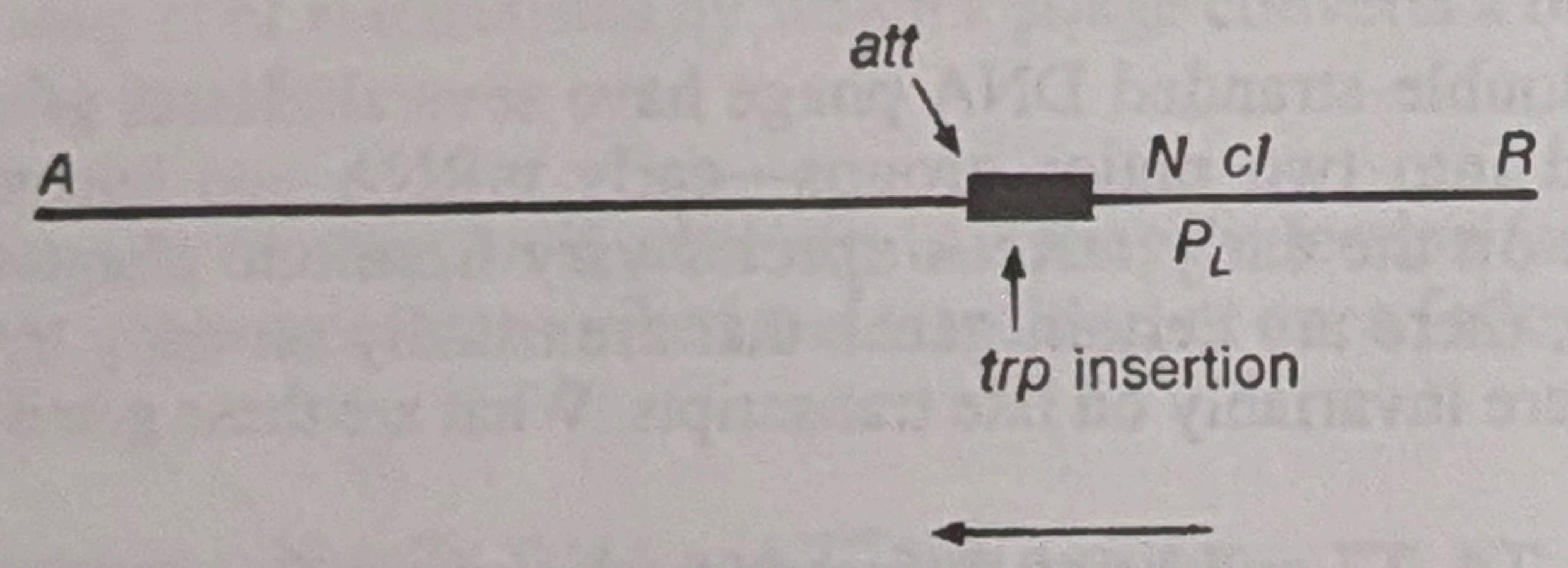
28 What functions are lacking in a λ Q^- mutant?

29 A deletion of λ called *nin5* removes a large segment of DNA between genes *P*

and *Q*. This deletion allows growth of a λ phage that is mutant in an important early gene. Which gene is it?

30 Suppose a λ phage has infected *E. coli* and is actively replicating its DNA. No DNA has been packaged, since late mRNA has not yet been made. If a T4 phage is then added, will the λ phage still be made?

31 There exist certain λ particles that carry the genes for synthesizing tryptophan (*trp*-transducing particles). The locus of the *trp*-gene insertion is shown in the figure below. The *trp* operon is transcribed in the direction indicated by the horizontal arrow, that is, in the same direction as the phage mRNA that goes leftward from *pL*. There are several classes of *trp*-transducing particles that differ according to the size of the λ DNA replaced by *E. coli trp* DNA. In each class, the size of the inserted bacterial DNA is the same as the DNA that is absent, the bacterial DNA contains the tryptophan synthetase gene, the insertion begins at *att* and moves to the right, and no essential phage genes are missing. When these particles infect a bacterium lacking tryptophan synthetase, this enzyme is made. (a) If the phage carries a point mutation in the λ *N* gene, no tryptophan synthetase is made. Why not? What part of the *trp* operon must be missing ? (b) Consider a *trp* particle whose bacterial DNA segment is so large that part of the *N* gene is replaced by bacterial DNA. Will any tryptophan enzymes be made? Explain briefly.



32 How many λ particles can be packaged from a dimer circle? A trimer circle?

33 If λ P^- infects *E. coli* at a low moi, no phage are produced because there is no DNA replication. If moi = 10 and either the bacterial or phage recombination systems are active, about one-third of the infected cells release one phage. Explain.

34 (a) For *E. coli* phage ϕX174, would you expect the (+) and (–) circular, single-stranded DNA strands of the phage (isolated from the double-stranded replicative form) each to be infective in transfection experiments? (b) Would you also expect the (+) and (–) RNA strands of the *E. coli* RNA phages to be individually infective?

35 (a) How does ϕX174 manage to include all of its necessary genetic information in such a small DNA molecule? (b) Do you think that ϕX174 could reproduce in a heavily UV-irradiated host?

36 Which phage would be expected to produce clearer plaques, ϕX174 or M13? Why?

ANSWERS TO PROBLEMS

1 The *mal* operon is inducible, so growth in maltose medium increases the number of maltose-transport systems. Inhibition by glucose suggests that the *mal* operon requires bound cAMP-CRP and hence is regulated by glucose.

2 Clearly, the receptors never evolved in the first place in order that the bacterium could be infected. Thus, the receptors must be important to the bacterium. For example, the receptor for phage T1 is a transport system for the Fe^{2+} ion. Receptor sites might gradually change in time to reduce the accessibility of certain phages but the change will be limited by the number of amino acids that can be altered without loss of essential activity.

3 The bacteria had stopped growing when the phage were added, so phage growth is not possible.

4 (a) Small plaques will usually result if the burst size is especially small or if the phage adsorbs very slowly to bacteria. If the phage particle is very small, the plaque will usually be large because smaller phages diffuse further in agar. (b) If a phage is adsorbed to a bacterium before dilution into agar, new phage will be produced sooner because the phage does not have to find a bacterium in the agar. Diffusion of phage in agar is sufficiently slow, especially for the tailed phages, that one or two cell generations might elapse before all phages manage to adsorb. Since there is only a limited time in agar in which phage can be synthesized (until bacterial growth stops), the size of the plaque depends upon how quickly a phage particle adsorbs to a bacterium.

5 T7 has no tail and diffuses more rapidly from the site of initial infection than does T4. Thus, soon after the second cycle of infection, T7 phage particles are quite far (on a microscopic scale) from the cell that was initially infected.

6 T4 will be turbid, because half of the bacteria will never be infected; that is, B/4 can continue to grow in the plaque. T4*h* will be clear, because all bacteria will be infected.

7 The cell culture lyses visibly at 60 minutes. This time is 20 minutes after the moi becomes greater than 1 phage per cell. The calculation is the following: After 20

min there are (200)(5000) = 10^6 phage/ml. At t = 30 min the bacteria have doubled to ca. 4 x 10^7 per ml. At t = 40 the number of phage per ml is 200 x 10^6 = 2 x 10^8, which is greater than the number of bacteria. Thus, lysis will occur one phage life cycle, or 20 minutes later, at t = 60. 8. After adsorption, divide the mixture in half. Add antiserum to one half to kill unadsorbed phage. Measure the number of infected cells by their ability to produce phage; this is the number of infective centers. To the other half, add chloroform to kill infected cells and measure unadsorbed phage. The moi equals (phage added – unadsorbed phage)/(infected cells).

9 From Poisson's law e^{-2} or 0.135 of the bacteria are uninfected. Therefore the number of plaques is 200(1 – 0.135) = 173.

10 A phage with a burst size of 2 could not form a plaque unless it had an exceedingly short life cycle, such as 5 minutes.

11 (a) These are bacteria to which no phage adsorb. Since the total moi = 5, the fraction of cells which is uninfected is e^{-5} = 0.007. The number of uninfected cell is 7 x 10^5. (b) This number includes bacteria to which P4, but not P2, adsorb, as well as the infected cells. The moi of P2 is 3; the fraction of cells getting no P2 is e^{-3} = 0.05. Since 0.007 get neither P2 nor P4, then 0.05 – 0.007 = 0.043 get one or more P4 but no P2. Hence 0.043 x 10^8 = 4.3 x 10^6 bacteria get P4 but no P2. (c) These are the bacteria to which P2 adsorb but no P4 adsorb. Using the above reasoning, e^{-2} = 0.135 get no P4, and 0.007 get neither; hence 0.135 – 0.007 = 0.128, and 1.28 x 10^7 cells adsorb P2 but no P4. (d) These are the bacteria infected by at least one P2 and one P4. The number of cells getting at least one P4 and one P2 is 10^8 – (1.28 x 10^7) – (4.3 x 10^6) – (7 x 10^5) = 8.22 x 10^7

12 By destruction of host DNA; by conversion of host RNA polymerase to a form which cannot read host promoters; by inactivation of host RNA polymerase and replacement with phage-coded polymerase.

13 Lysis would probably occur before many (or any) phage were made.

14 T4 DNA is cyclically permuted, it contains 5-hydroxymethylcytosine instead of cytosine, it is glucosylated, and it is 4.5 times larger than T7 DNA. Both are terminally redundant.

15 T4 uses the host RNA polymerase throughout the life cycle, but modifies it several times. T7 starts with the host RNA polymerase and then inactivates it and uses a phage-encoded RNA polymerase. Both use a phage-encoded DNA polymerase.

16 In a simultaneous infection the T4 nuclease destroys the T7 DNA. If T4 infection occurs 10 minutes after T7 infection, the T4 DNA is not transcribed because T7 has inhibited the *E. coli* RNA polymerase. Therefore, T7, but not T4 phage, are produced.

17 To prevent incorporation of cytosine into T4 DNA.

18 The T4-induced nucleases will destroy all newly synthesized phage DNA molecules since, in cases (a) and (b), glucosylation cannot occur.

19 No. The DNA is terminally redundant, so the injection sequence differs for each particle.

20 Modify *E. coli* RNA polymerase (1) to prevent recognition of promoters that have been used and are no longer needed and (2) to permit recognition of the next promoter that is needed.

21 (1), (3), and (4).

22 Yes. If the headful rule is obeyed, the deletion will produce terminal redundancy.

23 The early transcripts always contain genes for DNA replication and genes that (in varied ways) regulate transcription. Late mRNA generally encodes the structural protein genes and the lysis system. T7 is an exception in that the genes encoding replication functions and the structural proteins are initially on a single transcript.

24 T4 successively modifies RNA polymerase to recognize new promoters; T7 makes its own RNA polymerase, inactivates the *E. coli* enzyme, and uses slow injection to delay late transcription; λ alters the ability to ignore transcription termination signals.

25 The structural proteins from the late transcripts are needed in stoichiometric amounts. The early proteins are mostly enzymes needed in tiny amounts. Since both early and late mRNA molecules have nearly the same half-lives, more copies of late mRNA are needed than early mRNA.

26 DNA ligase, DNA gyrase.

27 L1, R1, R2, and R4.

28 No late mRNA is made. Therefore, there are no structural proteins, DNA concatemers are not cleaved, and the cells do not lyse.

29 Gene *N* can be mutant, because the deletion eliminates the termination site located before gene *Q*. The symbol *nin* stands for *N*-*in*dependence.

30 No, because T4 can be transcribed and the T4-encoded nucleases will destroy the λ DNA.

31 (a) Messenger RNA terminates at the first leftward terminator *tL1* of λ, and N protein is needed to prevent termination. This conclusion is correct as long as the *trp* operon lacks a promoter. (b) Yes. If part of gene *N* is deleted, *tL1* must also be deleted and the mRNA from the main leftward promoter will extend through the *trp* operon.

32 For the dimer, after the first DNA molecule is cut out and packaged, the remaining DNA has no *cos* site and hence cannot be packaged. Thus, only one particle can be formed from a dimeric circle. By the same reasoning, two particles are packaged from a trimeric circle.

33 Recombination produces a circular dimer in some of the infected cells.

34 The point of this question is that if a phage codes for its own replicase or initiator for replication, then a successful infection can occur only if the nucleic acid in the phage particle is the coding strand. Also, if the replication origin were on a particular strand, then only that strand would be infectious. (a) In φX174 the origin is on the (+) strand so the (–) strand is not infective. (b) In the RNA phages the (+) strand encodes the replicase, so the (–) strand is not infective.

35 (a) It has overlapping genes. (b) No, for it has so few genes that it surely must use host functions.

36 Only φX174 would make clear plaques because M13 does not lyse its host.

MORE CHALLENGING PROBLEMS

1 The number of phage in a lysate is being determined by plating. First, 0.1 ml of the lysate is diluted in 10 ml of buffer (tube A). Two successive dilutions of 0.1 ml to 10 ml are made (tubes B and C), followed by two successive dilutions of 1 ml to 9 ml (tubes D and E). Then, 0.1 ml from various dilution tubes is placed in agar with sensitive bacteria, and the next day, plaques are counted. The number of plaques obtained from the 0.1 ml aliquots from tubes C, D, and E are 2010, 352, and 18, respectively. What is the best estimate of the number of phage per ml in the lysate? Why do the number of plaques per dilution tube not simply follow the dilution?

2 Consider a phage with the strange property that an infected cell will yield progeny phage only if at least two phage adsorb to the bacterium. If you have 10^8 bacteria and add to this culture 3×10^8 phage, how many bacteria will produce phage?

3 A new protein X appears in infected cells. Describe various experiments that you might perform to prove that the gene coding for X is phage-encoded and is not encoded in host DNA.

4 How would you show whether or not a phage-encoded gene product is required throughout the infectious cycle or only at a unique time?

5 The genes *J* and *E* of phage λ encode a tail protein and the major head protein respectively. The E^- and J^- mutations are amber mutations and the bacteria used in the following experiments do not contain a suppressor. If *E. coli* is infected with an E^- phage, no viable phage are produced but lysis occurs. Such a lysate is called an E^- lysate. A similar infection with λ J^- yields a J^- lysate, which also contains no viable phage. If E^- and J^- lysates are mixed, viable phage form in numbers reaching about 50 phage per original cell. To study this phenomenon the following procedure is carried out. The E^- lysate is mixed with *E. coli* which has an amber suppressor, and the bacteria are then removed by centrifugation. The supernatant is devoid of activity; that is, complementation with the J^- lysate is abolished. However, the sediment, which contains all of the added cells, is active, in the sense that if the J^- lysate is added, the cells develop phage and lyse. (a) What is happening in this experiment? (b) What is the genotype of the phage produced by the infective centers?

ANSWERS TO MORE CHALLENGING PROBLEMS

1 3.5×10^{10}, which is calculated from tube D. The number in tube E is small and statistically less reliable. The numbers fail to follow dilution, because the number in tube E is statistically unreliable; that is, the number in tube C is so large that many plaques must overlap and two plaques will be counted as one.

2 Use Poisson's law. The fraction of cells infected by two or more phage $= 1 - P(0) - P(1) = 0.8$. Therefore 80% of 10^8 bacteria ($= 8 \times 10^7$ bacteria) are infected by two or more phage.

3 There are several approaches, each of which involve isolating phage mutants that eliminate the activity of X. It is also necessary to have an *in vitro* assay for X; this might be enzymatic or immunological, or might require isolation of the protein. First, consider a heat-sensitive mutant. With this, it can be shown that at high temperature, enzymatic activity or perhaps immunological reactivity (not as good a test) is substantially reduced. Second, consider a chain termination mutation. Here it can be shown by sedimentation or by gel electrophoresis that the protein has a lower molecular weight than the wild-type protein. Note the greater power of the second approach. With the heat-sensitive mutant, one

30 No, because T4 can be transcribed and the T4-encoded nucleases will destroy the λ DNA.

31 (a) Messenger RNA terminates at the first leftward terminator *tL1* of λ, and N protein is needed to prevent termination. This conclusion is correct as long as the *trp* operon lacks a promoter. (b) Yes. If part of gene *N* is deleted, *tL1* must also be deleted and the mRNA from the main leftward promoter will extend through the *trp* operon.

32 For the dimer, after the first DNA molecule is cut out and packaged, the remaining DNA has no *cos* site and hence cannot be packaged. Thus, only one particle can be formed from a dimeric circle. By the same reasoning, two particles are packaged from a trimeric circle.

33 Recombination produces a circular dimer in some of the infected cells.

34 The point of this question is that if a phage codes for its own replicase or initiator for replication, then a successful infection can occur only if the nucleic acid in the phage particle is the coding strand. Also, if the replication origin were on a particular strand, then only that strand would be infectious. (a) In ϕX174 the origin is on the (+) strand so the (–) strand is not infective. (b) In the RNA phages the (+) strand encodes the replicase, so the (–) strand is not infective.

35 (a) It has overlapping genes. (b) No, for it has so few genes that it surely must use host functions.

36 Only ϕX174 would make clear plaques because M13 does not lyse its host.

MORE CHALLENGING PROBLEMS

1 The number of phage in a lysate is being determined by plating. First, 0.1 ml of the lysate is diluted in 10 ml of buffer (tube A). Two successive dilutions of 0.1 ml to 10 ml are made (tubes B and C), followed by two successive dilutions of 1 ml to 9 ml (tubes D and E). Then, 0.1 ml from various dilution tubes is placed in agar with sensitive bacteria, and the next day, plaques are counted. The number of plaques obtained from the 0.1 ml aliquots from tubes C, D, and E are 2010, 352, and 18, respectively. What is the best estimate of the number of phage per ml in the lysate? Why do the number of plaques per dilution tube not simply follow the dilution?

2 Consider a phage with the strange property that an infected cell will yield progeny phage only if at least two phage adsorb to the bacterium. If you have 10^8 bacteria and add to this culture 3×10^8 phage, how many bacteria will produce phage?

3 A new protein X appears in infected cells. Describe various experiments that you might perform to prove that the gene coding for X is phage-encoded and is not encoded in host DNA.

4 How would you show whether or not a phage-encoded gene product is required throughout the infectious cycle or only at a unique time?

5 The genes *J* and *E* of phage λ encode a tail protein and the major head protein respectively. The E^- and J^- mutations are amber mutations and the bacteria used in the following experiments do not contain a suppressor. If *E. coli* is infected with an E^- phage, no viable phage are produced but lysis occurs. Such a lysate is called an E^- lysate. A similar infection with λJ^- yields a J^- lysate, which also contains no viable phage. If E^- and J^- lysates are mixed, viable phage form in numbers reaching about 50 phage per original cell. To study this phenomenon the following procedure is carried out. The E^- lysate is mixed with *E. coli* which has an amber suppressor, and the bacteria are then removed by centrifugation. The supernatant is devoid of activity; that is, complementation with the J^- lysate is abolished. However, the sediment, which contains all of the added cells, is active, in the sense that if the J^- lysate is added, the cells develop phage and lyse. (a) What is happening in this experiment? (b) What is the genotype of the phage produced by the infective centers?

ANSWERS TO MORE CHALLENGING PROBLEMS

1 3.5×10^{10}, which is calculated from tube D. The number in tube E is small and statistically less reliable. The numbers fail to follow dilution, because the number in tube E is statistically unreliable; that is, the number in tube C is so large that many plaques must overlap and two plaques will be counted as one.

2 Use Poisson's law. The fraction of cells infected by two or more phage $= 1 - P(0) - P(1) = 0.8$. Therefore 80% of 10^8 bacteria ($= 8 \times 10^7$ bacteria) are infected by two or more phage.

3 There are several approaches, each of which involve isolating phage mutants that eliminate the activity of X. It is also necessary to have an *in vitro* assay for X; this might be enzymatic or immunological, or might require isolation of the protein. First, consider a heat-sensitive mutant. With this, it can be shown that at high temperature, enzymatic activity or perhaps immunological reactivity (not as good a test) is substantially reduced. Second, consider a chain termination mutation. Here it can be shown by sedimentation or by gel electrophoresis that the protein has a lower molecular weight than the wild-type protein. Note the greater power of the second approach. With the heat-sensitive mutant, one

merely correlates a property *in vitro* with that *in vivo*, but with a chain-termination mutant one observes not only the absence of the wild-type protein but the appearance of a new and smaller molecule.

4 The best way is to use a temperature-sensitive mutant and raise the temperature of the infected cell at various times. If the gene product is needed only in the interval between t_1 and t_2, then raising the temperature later than t_2 will not affect phage development and a period of heating at any time between zero time and $t = t_1$, followed by cooling, will usually not affect phage development.

5 (a) The E^- lysate contains active tails. These tails can attach to live *E. coli*. Filled heads from the J^- lysate attach to the preadsorbed tails. Following attachment, the λJ^- DNA can inject and the infection can proceed, producing J^- progeny. (b) The genotype is J^- because this is the genotype of the DNA contained in the tail-less heads.

CHAPTER 18

BACTERIOPHAGES II: THE LYSOGENIC LIFE CYCLE

CHAPTER SUMMARY

In the life cycle of a virulent phage a bacterium is infected and progeny phage are produced. Except for phages in the M13 class, the bacterium is killed and lyses in the process. Temperate phages can enter either a lytic or a lysogenic life cycle. In a lysogenic cycle early genes are initially transcribed, but late mRNA is not made. No progeny phage are made and ultimately all expression of the phage genes is repressed, except for synthesis of the phage repressor. Phage DNA is ultimately inserted into the bacterial chromosome. The bacteria grow and divide, as if the phage DNA were not present, and the phage DNA is carried from one bacterial generation to the next as part of the bacterial genome. The lysogenic cycle occurs when the multiplicity of infection is high and the bacterial culture is begin to enter the stationary phase. A bacterium containing integrated phage DNA is called a lysogen, the integrated phage DNA is a prophage, and the overall phenomenon is lysogeny. The integrated prophage DNA is stable, but if the bacterial DNA is damaged, derepression (prophage induction) occurs, the phage DNA is excised, and phage development occurs.

The insertion process is a recombinational exchange in which phage and bacterial DNA recombine within particular nucleotide sequences called attachment sites. Distinct phage and bacterial attachment sites, *POP'* and *BOB'*, respectively, possess a common 15-bp core sequence in which the exchange occurs. All temperate phages circularize their DNA after infection. Since the phage has a single attachment site, the order of the prophage genes is a permutation of the order of the genes in the phage particle. Insertion is catalyzed by a phage-encoded protein called integrase, or Int, which is aided by a bacterial protein—integration host factor. Integration generates two prophage attachment sites—*BOP'* and *POB'*. Excision involves recombination between the prophage attachment sites. However, integrase plus the integration host factor cannot catalyze an exchange between *BOP'* and *POB'*; this exchange requires a phage-encoded protein called excisionase, or Xis. During integration, synthesis of excisionase in prevented by degradation of any mRNA molecule that includes

an intact *xis* gene and by activation of a new promoter for Int within the *xis* gene. The *int* promoter is activated by the cII protein, which simultaneously activates synthesis of the cI repressor.

There are two promoters for the λ cI repressor. One, *pre*, is used for establishment of synthesis of the repressor during the early stages of the lysogenic cycle. It does not remain active for very long. The other, *prm*, is activated by the cI repressor and maintains repressor synthesis, not only during the remainder of the lysogenic cycle but also in subsequent bacterial generations. Thus, all lysogenic bacterial maintain a constant concentration of cI repressor. If a λ lysogen is infected with another λ phage, the intracellular repressor binds to two λ operators and thereby prevents expression of genes of the incoming DNA. Thus, λ can neither multiply nor form plaques on a λ lysogen. This phenomenon is called immunity. A phage with a different repressor-operator system, for example, phages 21 and 434, can plate on a λ lysogen. Hybrid phages consisting of λ DNA with the repressor-operator segment of either phage 434 or 21 have been prepared. These so-called heteroimmune phages can plate on a λ lysogen.

The excision process is normally accurate, but at low frequency aberrant excision occurs, giving rise to transducing particles, which lack λ genes from one end of the prophage and include bacterial genes from the host DNA at the opposite end of the prophage. Transducing particles can infect and produce progeny if only nonessential genes have been lost. However, both viable and nonviable particles can transfer the bacterial genes to a second host. By recombination these genes can become integrated into the host DNA. Transducing particles formed by aberrant excision can carry only bacterial genes that flank the prophage and hence are called specialized transducing particles.

Some phages, for example, *E. coli* phage P1, possess a repressor-operator system and lysogenize without integration into the host DNA. Instead, the repressed DNA forms a plasmid, which replicates once per bacterial generation, thereby maintaining itself throughout subsequent bacterial generations. How this replication is regulated is unknown. P1 produces transducing particles but of a different type from the lambdoid phages. When P1 infects a bacterium, the host DNA is fragmented, as in a T4 infection. However, the P1 DNA-packaging system is not fastidious with respect to DNA type or size and occasionally packages fragments of host DNA into a phage particle. These transducing particles, which contain no phage DNA, can inject donor DNA into a recipient bacterium; by a recombination mechanism the transferred DNA can replace homologous DNA of the recipient, generating a recombinant bacterium called a transductant. P1 is called a generalized transducing phage because any bacterial gene can be transduced by it.

BOLD TERMS

core sequence, deletion-substitution particles, excisionase, generalized transducing phages, heterogenotes, heteroimmune, homoimmune, immunity, induction, insertion, integrase, integrated, integration, lysogen, lysogenization, prophage, prophage attachment sites, prophage induction, resistance to homoimmune superinfection, specialized transducing particle, specialized transducing phage, temperate, transducing particle, transducing phage, transduction, virulent

ADDITIONAL INFORMATION

In the text, it was stated without proof that λ prophage is linearly inserted between the *gal* and *bio* genes. Evidence for this conclusion came from both genetic and physical experiments. The initial genetic experiment was that described in the text, namely, that the gene order of the prophage was permuted with respect to the gene order of the phage. Other experiments utilized generalized transduction with phage P1. In these experiments it was shown that the *gal* gene cotransduced with prophage genes at the left end of the prophage map and that the *bio* gene cotransduced with phage genes at the right end of the prophage map. Furthermore, *gal* and *bio* could be cotransduced from a nonlysogen, but not from a lysogen; in a lysogen the genes are too far apart to fit in a P1 head. The physical experiment utilized a circular DNA plasmid that contained *BOB'*. This plasmid was placed in a cell whose *BOB'* had been deleted. Thus, when the cell was lysogenized, insertion could only occur into the plasmid. Molecular weight measurements showed that the molecular weight of the plasmid (105 million) increased to 136 million when lysogenization occurred, the difference being precisely the molecular weight of λ DNA (31 million); furthermore, both the lysogenic and nonlysogenic plasmid molecules were circular.

Whereas λ is certainly the prototype for a temperate phage and integration of most phages follows the λ pattern, both qualitative and quantitative differences have been observed with other phages. For example, insertion of λ occurs primarily at a single site in the bacterial chromosome, between the *gal* and *bio* genes. However, some phages insert their DNA within genes and thereby cause mutations, and other phages have more than one major insertion site. The particular insertion site can be identified by Hfr mapping or by determining the type of specialized transducing particles produced by an induced lysogen. One curious phage has its normal attachment site within its *int* gene. When integration occurs, the *int* gene is interrupted; hence, excision of the prophage can never occur!

λ can also insert in secondary sites. The frequency is very low and such insertions are observed only when a bacterial mutant is used in which the λ attachment site is deleted. Integration is detected by the production of immune cells. Excision from secondary sites is also infrequent and is aberrant, producing transducing particles. This technique of lysogenizing a deletion mutant has enabled geneticists to obtain a large number of transducing particles containing bacterial DNA from many parts of the chromosome.

The integrase protein also mediates site-specific exchange between phage DNA molecules. λ contains the attachment site *POP′*, and integrase is capable of catalyzing an exchange between two *POP′* sites. Thus, in the absence of all standard recombination systems λ is still able to recombine, but only genes on opposite sides of the attachment site can recombine. For example, a J^-P^+ and a J^+P^- phage can recombine and produce J^+P^+ and J^-P^- recombinants because genes *J* and *P* are on opposite sides of *att*. However, no recombination can occur between genes *F* and *J*, which are on the same side of *att* . This mode of phage-phage recombination, which is known as Int-promoted recombination, has been used to study the mode of prophage integration and excision. All that is needed is a pair of phages carrying different attachment sites and a mutation in the gene determining normal phage-phage recombination. Experiments are carried out in bacteria defective in bacterial recombination, so that the only recombination that can occur is Int-promoted recombination. Gal^- and bio^- transducing phages do not have the attachment site *POP′*, but instead, because of the aberrant excision, have the attachment sites *BOP′* (*gal*) and *POB′* (*bio*): thus, a cross between a *gal* and a *bio* phage, namely, *BOP′* x *POB′*, is precisely the exchange that occurs in prophage excision. The product of the cross is a phage carrying *POP′* and another with *BOB′*. A cross between these phages, *POP′* x *BOB′*, mimics prophage integration. The original *in vitro* experiment in which integrase was first detected used a curious phage mutant carrying both *att* sites *BOB′* and *POP′* flanking a segment of bacterial DNA. A circular DNA molecule was used in the assay and Int-promoted recombination was detected by the excision of a DNA fragment between the two *att* sites producing a smaller circle of solely bacterial DNA.

Lysogeny has a practical use in a test, which, like the Ames test (Chapter 10), can identify cancer-causing chemicals (carcinogens). Most carcinogens are, for reasons that are not yet understood, able to induce a lysogen to produce phage. Lysogens are spontaneously induced at a fairly low frequency and in certain bacterial mutants the frequency is lower. Spontaneous induction can be detected quite simply. If one lysogen is mixed with 10^8 bacteria and put on solid medium (exactly as might be done in detecting phage by plaque counting), the lysogen, which is not induced by this process, forms a microcolony (of at most a few hundred cells) in the bacterial lawn. Occasionally, late in the development of the colony, a cell will be spontaneously induced, but this does not usually produce a plaque because depletion of nutrients in the growth medium limits cell growth, so the released phage cannot multiply (remember: phage multiply only in growing cells). However, if an inducing agent (in this case, a carcinogen) is in the agar, the original lysogenic cell will be induced to make phage at the time all of the bacteria in the agar begin to grow, so a plaque will form. If 10^5 lysogens are mixed with the indicator bacteria, the effectiveness of the inducing agent can be measured by the fraction of lysogenic cells yielding plaques. This test, known as the Devoret test, is a sensitive means of detecting carcinogens.

DRILL QUESTIONS

1 How does a temperate phage differ from a virulent phage?

2 Which phage can plate on a λ lysogen: λ*imm*λ, λ*imm*434, λ*imm*21?

3 Which phage can plate on a lysogen of λ*imm*434: λ*imm*λ, λ*imm*21, λ*imm*434?

4 Does lysogeny occur at low or high multiplicity of infection?

5 At what stage of the growth cycle of a bacterial culture will lysogeny occur at high frequency, if the phage multiplicity of infection is correct?

6 Two temperate phages are homoimmune. What phage elements must they have in common?

7 What is meant by a virulent mutant of a phage?

8 What plaque morphology is produced by a temperate phage?

9 What plaque morphology is produced by a cI^- mutant and by a *vir* mutant?

10 What do you think is the difference, if any, between the appearance of a plaque of a *vir* phage on a homoimmune lysogen or a nonlysogen?

11 Phage 434 does not insert between the *gal* and *bio* genes, because its attachment site is at another position. At what locus in the chromosome would you expect λ*imm*434 to insert if only the promoter-operator region of phage 434 were present in the hybrid phage?

12 What symbol designates the bacterial and phage *att* sites (i.e., *B*, *P*, etc.)?

13 What symbols designate the prophage attachment sites and which is at the *gal* and at the *bio* end?

14 A phage has gene order *A B C att D E F*. What is the gene order in a prophage?

15 What two λ proteins are regulated by the cII protein?

16 What proteins are needed for integration?

17 What proteins are needed for excision?

18 What gene products activate the *pre* and *prm* promoters for cI synthesis?

19 What attachment sites are carried by *gal*- and *bio*-transducing particles?

20 If 10^8 phage λ infect 10^8 gal^+ bacteria and all phage adsorb, what fraction of the infected cells will yield *gal*-transducing phage?

21 Remember that a λ prophage is located between the *E. coli gal* and *bio* genes. The *lac* gene is about 10 minutes away from the *gal* gene in the *E. coli* time-of-entry map. Roughly what would be the ratio of *lac*- and *gal*-transducing particles?

ANSWERS TO DRILL QUESTIONS

1 A temperate phage may choose between a lytic and lysogenic cycle; a virulent phage can only engage in a lytic cycle.

2 λ*imm*434 and λ*imm*21.

3 λ*imm*λ, λ*imm*21.

4 High.

5 Stationary phase.

6 Generally their repressors and repressor-binding sites (operators) must be the same, though it is sufficient if the repressor of each phage can recognize the operator of the other.

7 A phage that can plate on a homoimmune lysogen.

8 The plaque has a turbid center surrounded by a clear area.

9 Both make clear plaques.

10 No difference.

11 At the normal bacterial attachment site, since the hybrid phage has λ *att*.

12 Bacterial, *BOB'*; phage, *POP'*.

13 *Gal, BOP'; bio, POB'*.

14 *DEFABC*.

15 Int and cI.

16 The phage integrase and the bacterial integration host factor.

17 The phage Int and Xis proteins.

18 *Pre* is activated by the cII protein; *prm* is activated by the cI protein.

19 *Gal, BOP'; bio, POB'*.

20 None, because specialized transducing particles are produced only by lysogens.

21 There will be no *lac*-transducing particles, because the *lac* locus is too far away to be included in a λ phage head.

PROBLEMS

1 The *cI* gene of λ codes for the immunity repressor. If the λ is cI^-, the plaque is uniformly transparent. If it is cI^+, the plaque has a turbid center. Explain the plaque morphology. Would you expect T2 plaques ever to have turbid centers?

2 Mutations in what genes or sites would prevent lysogenization?

3 An int^- mutant can integrate if the cell is coinfected with an int^+ phage, the latter "helper" phage provides the missing function. Could a phage with an operator mutation (e.g., *vir*) be helped to integrate?

4 A normal λ phage cannot successfully infect a bacterium lysogenic for normal λ (that is, it cannot make more phage and lyse the bacterium). (a) Why not? (b) Suppose the phage added to the lysogen is a mutant in that the left (*oL*) and the right (*oR*) operators are mutated (that is, they are defective in the same way that a *lac* operon o^c mutant is defective). Can this phage successfully infect a wild-type lysogen? Explain briefly.

5 If one λ phage particle infects a cell lysogenic for λ, which of the following will probably happen: (1) a normal phage cycle producing about 50 phage particles will occur; (2) the DNA will circularize but will not replicate; (3) the cell will die; (4) the prophage will be excised; (5) the DNA will not be injected?

6 A temperate phage has the gene order *A B C D E F G H* and a prophage gene order *G H A B C D E F*. What information does this give you about the phage?

7 What genetic property would a phage have if the gene order of the prophage were identical to that in the free phage?

8 A λ *gal* phage can lysogenize a heteroimmune lysogen to form a tandem dilysogen. In theory, it could insert into the left (*gal*) or right (*bio*) prophage attachment sites. Genetic tests show that the gene order of the dilysogen is *gal* *BOP'* *R A gal* *BOP'* *R A* *POB'* *bio*, in which *A* and *R* are genes at the termini of the DNA (given for reference only). Into which prophage *att* site did insertion occur?

9 A λ phage genetic map (as obtained in standard crosses) containing only a few of the known genes is approximated below. Which of these genes would show the highest frequency of cotransduction with the *gal* gene if a P1 phage grown in a gal^+ λ-lysogenic cell were used to transduce a gal^-(λ) lysogen?

A B J cI P O R

10 You have isolated two independent mutant strains of the phage λ; both of these mutants form clear plaques. When either mutant alone infects *E. coli* at moi = 10, no lysogenic survivors can be isolated. However, when *E. coli* is coinfected with 5 phage per cell of the first mutant plus 5 phage per cell of the second mutant, then 10% of the cells survive the infection, and almost all of these survivors are lysogenic. When the lysogenic survivors are induced, 99 of the plaques produced are clear. (a) What λ function or functions do you think are eliminated in these mutants? (b) What is the most likely explanation for the increased frequency of lysogenization which occurs during coinfection by the two mutants?

11 Which of the following λ mutants cannot lysogenize at normal frequency: N^-, P^- Q^-?

12 Phage variants have been prepared containing many different attachment sites. (a) What attachment sites are in *bio*- and *gal*-transducing phages? (b) What attachment sites are generated by Int-mediated recombination between these phages?

13 Could a transducing particle carry *POP'*?

14 Genetic recombination between λ phages can occur by a variety of systems. Two of these, the bacterial Rec system and the phage Red system, can be eliminated by mutation. However, recombinants still arise in a mixed infection with genetically marked phages, but only between certain markers. For example, recombination occurs between the following pairs of genes: *A–R*, *A–O*, *B–O*, and *B–R*. However, no recombinants can be detected in crosses between *A* and *B*, and between *O* and *R*. Explain these results.

15 Could an A^+R^+ recombinant be produced in a cross between λA^-R^+bio and A^+R^-bio, if both the Rec and Red systems are inactive? Explain. The A and R genes are at opposite ends of the DNA molecule.

16 If a lysogen is irradiated with UV light and the cells are plated with an indicator bacterium, plaques form, roughly one per irradiated cell. If the cell is *recA*–, no plaques are formed. If the cell contains a temperature-sensitive cI repressor, phage are produced with both $recA^+$ and $recA^-$ bacteria. Explain this observation.

17 Remember the *E. coli* gene order *gal att bio*, in which *att* = *BOB'* is the attachment site in a nonlysogenic bacterium. If phage P1 is grown on a nonlysogen, some of the transducing particles carry the *gal* gene, some carry *bio*, and some carry both *gal* and *bio* genes. The ratio of particles transducing both *gal* and *bio* together to those transducing gal is 0.01. If the P1 phage is grown instead on a λ lysogen, would you expect the number of particles transducing the *gal* and *bio* loci together to increase, decrease, or stay the same? Why?

18 Infection of a λ lysogen with 10 λ phage particles does not result in phage development. This is the immune response. If the multiplicity of infection is 30, phage development occurs. In a dilysogen (carrying two prophages) 60 phage particles per cell are needed to initiate a successful infection. Explain.

19 The temperature-sensitive λ*cI* mutant, *cI857*, is especially useful, because a lysogen need only be heated to 40 C to induce phage development, and induction is a convenient way to grow the phage. The *cI857* repressor is reversibly denaturable in that if, after heating, a cell is returned to 34 C or less, the repressor renatures and becomes active. (a) At 40 C is the *cI* gene of a lysogen still transcribed? (b) If the temperature is returned to 34 C, will the *cI* gene be transcribed? (c) If a growing *cI857* lysogen is heated for 10 minutes and then cooled, phage development occurs on continued incubation in growth medium. If the lysogen is instead cooled after 5 minutes, there is no phage development and the bacteria survive. Explain. (d) In (c) when there is no phage development, if the bacteria are plated to form colonies, the colonies are found to consist of nonlysogenic cells. Explain.

20 Suppose you have a tandem double lysogen (that is, two prophages are next to one another) and both prophages are int^- and are incapable of DNA replication. Also, both prophages contain a mutant repressor that is temperature-sensitive, so at high temperature mRNA can be made. If this double lysogen is heated to the high temperature, which of the following will occur? (1) No phage will be produced; (2) one phage will be produced per cell; (3) two phage will be produced per cell; (4) only transducing phage will be produced.

21 In an *E. coli* tandem dilysogen, the left and right prophages have the genotypes $cI857int^-P^-$ and $cI857int^-A^-$ respectively. The repressor of each is inactive at 42 C and the prophages lack a functional integrase. The mutations P^- and A^- are, in this experiment, just genotypic markers located as follows: *A F J att N P R*. If

the lysogen is heated to 42 C, what will be the genotype or genotypes of the phage progeny? Why will more phage particles be produced than in the experiment described in Problem 20?

22 What are the differences between a phage that produces only specialized transducing particles, one that produces only generalized transducing particles, and one that can produce both?

23 Why are λ specialized-transducing particles generated only by induction rather than by lytic infection?

24 Some temperate phages that integrate their DNA do not seem to form specialized transducing particles. Give several possible reasons for this observation.

25 Following are some facts about specialized transduction by phage P22 in *Salmonella typhimurium:* (1) Specialized transducing particles of P22 carrying the *proA* and *proB* genes can be generated. These genes are immediately adjacent to the prophage attachment site on the bacterial chromosome. The $proA^-$ and $proB^-$ specialized transducing particles have the behavior described below. (2) Lysates of the specialized transducing particles can go through the lytic cycle, but only in mixed infections. (3) In single infections the particles transduce by substitution; however, in mixed infections, the particles transduce by lysogenization. What is the molecular basis of each of the foregoing observations?

26 (a) Would you ever expect to find a λ transducing particle that can transduce genes further from the prophage than *gal* and *bio*? (b) A specialized transducing particle necessarily lacks phage genes. Do you think that a specialized transducing particle could form a plaque on normal bacteria?

27 Under what circumstances could a lysate of phage P1 containing transducing particles carry phage λ?

ANSWERS TO PROBLEMS

1 When a wild-type (cI^+) λ phage adsorbs to one bacterium, the lytic response ensues and many phage are produced that infect other bacteria. In time, the number of phage increases to the point that many bacteria are multiply infected. This favors the lysogenic response. Since the lysogens are immune to subsequent infection, they grow in the plaque and produce a turbid growth of bacteria. A cI^- phage cannot establish repression and, therefore, always enters a lytic cycle; hence, an infected cell can never survive to form a colony. Phage T2 lacks a repressor, so it always kills its host cell and, hence, would not be expected to produce a turbid plaque.

2 Mutations in the *cI, cII, cIII*, and *int* genes would prevent lysogenization. Since an active N product is needed to produce the cII and cIII products in appreciable quantities, *N* mutants will lysogenize quite poorly. Mutations in *att* will also prevent lysogenization. *Cro* mutants also have a greatly reduced lysogenization frequency, but the reasons are fairly complex.

3 No, because an operator mutation is cis-dominant.

4 (a) The repressor in the lysogen prevents transcription of the infecting phage DNA. (b) Yes, because they would not be recognized by the repressor.

5 Statement 2. Circularization is a physical process that does not need a phage gene product. Replication, however, requires prior transcription, which is inhibited by repression.

6 The prophage attachment site is between genes *F* and *G*.

7 The attachment site would be formed when the cohesive sites joined.

8 The left. If it inserted in the right, the order would be *gal* $\underline{BOP'}$ *R A* $\underline{POP'}$ *R A gal* $\underline{BOB'}$ *bio*.

9 When integrated into *E. coli*, the gene order is *gal cI P Q R A J bio*, so *cI* will show the highest cotransduction frequency with *gal*.

10 (a) Repression or establishment of lysogeny. (b) Complementation between two proteins was needed to establish repression.

11 Only N^-, because it fails to make the cII protein.

12 (a) λ*bio, POB'*; λ*gal, BOP'*. (b) λ*gal bio* has *BOB'*; λwild has *POP'*.

13 I don't see how. Maybe you can figure out some complicated way to make such a particle.

14 Genes on opposite side of *att* can be recombined by the activity of the integrase system, even when all other recombination systems are absent. This type of recombination is explained in the Additional Information section.

15 No, because the *bio* substitution eliminates the *int* gene. Int-promoted recombination is the only possible recombination mechanism that could carry out the exchange if the other recombination systems were absent.

16 Phage production is in each case a result of prophage induction. Induction by ultraviolet light cannot occur in a $recA^-$ cell, because the RecA product is

responsible for cleavage of the cI repressor. Induction occurs with a temperature-sensitive repressor in both mutant and wild-type since it is the high temperature that is inactivating the repressor.

17 Decrease, because the *gal* and *bio* genes are moved apart by the prophage.

18 If enough phage DNA molecules enter a bacterium, the intracellular repressor will be titrated.

19 (a) No, because in the absence of active repressor the *prm* promoter is inactive. (b) Yes, because the renatured repressor activates *prm*. (c) After 10 minutes enough DNA replication has occurred without repressor synthesis that there is insufficient repressor to bind all of the operators. After only 5 minutes of DNA replication, sufficient repressor is present. (d) Repression is reestablished, but the Int and Xis proteins have been made, so prophage excision occurs. The excised prophage is repressed though and cannot replicate, so bacteria free of prophage DNA arise by successive cell division.

20 (2), because the Ter system excises and packages the segment of DNA between the *cos* sites of the two prophages.

21 The genotype will be $cI857P^+A^+$, because the only mode of phage production is Ter-mediated excision at the two *cos* sites. More phage particles are produced than in Problem 20, because in Problem 21 one prophage is P^+ and DNA replication can occur. A typical number of phage is 12 per cell.

22 Specialized only—integrates into the chromosome and either does not cause host DNA fragmentation during phage development or cannot package host DNA fragments. Generalized only—fragments host DNA, can package host DNA fragments, and probably does not integrate. Both—integrates, fragments host chromosome, and can package host fragments.

23 They are only formed by aberrant excision of a prophage. In a lytic infection λ DNA is not inserted in the *E. coli* chromosome.

24 (1) Their excision system is perfect (unlikely). (2) When aberrant excision occurs, the fragment size is too large or too small to be packaged. (3) No known host genes are near enough to the prophage to be picked up.

25 At each end of the prophage map essential genes are present that are absent in both types of transducing particles. Thus, lytic growth is only possible when a helper phage providing the missing functions is present. In single infection, transduction occurs by substitution, because neither the right nor left prophage attachment sites can recombine with the bacterial attachment site. In mixed infection, either the transducing DNA and the helper phage DNA recombine or the attachment sites of the transducing DNA can recombine with a newly generated prophage attachment site.

26 (a) If a nonessential gene were deleted, the amount of bacterial DNA that could be included in a transducing particle would be greater than that from a normal-sized prophage. Hence, with such a deletion mutant, genes further from the prophage could be transduced. Alternatively, the bacterium could have a deletion either between the attachment site and the *gal* gene or to the left of the *gal* gene. Both possibilities would enable genes further than *gal* to be included in a transducing particle. (b) The ability to form a plaque depends on whether essential or nonessential genes are lacking in the transducing particle. With λ the *gal*-transducing particles always lack the genes for head and tail proteins and hence cannot form plaques. However, the *bio*-transducing particles lack nonessential genes such as *int*, *xis*, and *red* and are plaque-formers. Some *bio*-transducing particles have such a large *bio* substitution that the essential *N* gene is deleted; these particles do not form plaques. Interestingly, a lysogen was made in a bacterial strain in which there was a deletion between *att* and *gal*. This moved *gal* sufficiently near to the prophage that a *gal*-transducing particle was isolated in which only the nonessential *b2* region was removed, so the particle was a plaque-former.

27 If obtained by infecting a λ lysogen; the phage DNA would then be no different from any other bacterial DNA.

MORE CHALLENGING PROBLEMS

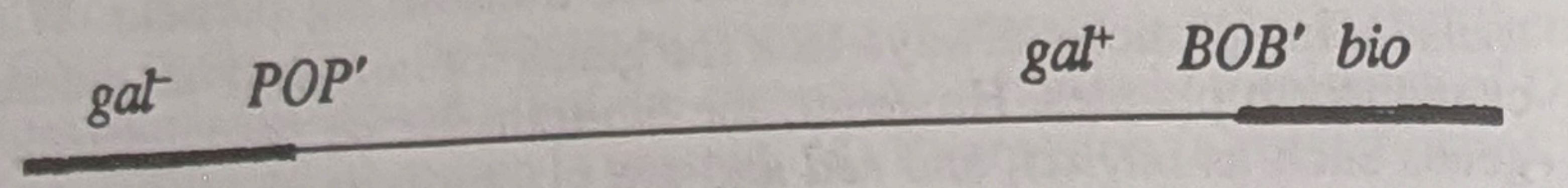
1 Suppose you have a λ mutant that makes a clear plaque on *E. coli* strain A and a turbid plaque on strain B. How can this difference be explained?

2 If a λ lysogen is induced with ultraviolet light, a lysate results in which about one clear-plaque mutant per 10^4 phage is present. If a culture of a lysogen is grown without any prior induction and the culture fluid is analyzed for free phage that are spontaneously released, the fraction that form clear plaques is always higher than 10^{-4}. (a) Explain the higher frequency of clear-plaque mutants among the spontaneously released phage. (b) Among the spontaneously released phage, in which of the genes *cI*, *cII*, or *cIII* do you suppose the mutations reside that yield a clear plaque?

3 You have given a friend a mutant λ phage that is defective in gene *J* and in one of the *red* genes needed for phage-mediated generalized recombination. She complains that growth of the J^-red^- phage in Rec^- cells lysogenic for a P^-red^- prophage yields J^+P^+ recombinants (the red^- mutations carried by the two phage types are identical). Can you explain this "extraordinary" recombination? How

might you show by a genetic experiment that your explanation is probably correct?

4 A λ mutant is isolated which is defective in that, although it forms plaques normally, a lysogen containing a single mutant prophage fails to produce phage when induced. The mutant lysogenizes with normal frequency. In studying this mutant, you infect bacterial strain R with this mutant at a moi = 3. Strain R has been obtained by lysogenizing a Gal$^-$ bacterium containing a segment of a prophage that has only the right prophage attachment site, with a *gal*$^+$ transducing phage. The resulting structure of this bacterium is the following:



Strain R contains a normal cI repressor and therefore is immune to superinfection. When R is infected with the λ mutant and the infected cells are plated on an indicator agar (galactose tetrazolium) in which Gal$^+$ and Gal$^-$ colonies are white and red, respectively, roughly equal numbers of white and red colonies are found. In what gene does the λ mutant map and what type of mutant is it?

5 A λ phage mutant called *b2* cannot lysogenize because it has a deletion starting just to the left of the *O* region of *att*; that is, its attachment site is Δ*OP′*, in which Δ represents a deletion. In a mixed infection with λ*bio* using conditions favoring the lysogenic response, dilysogens containing both λ*b2* and λ*bio* are found. Very few single lysogens (these can be ignored in your thinking) and no (*b2*,*b2*), or (*bio*,*bio*) dilysogens are found. How have the (*b2*, *bio*) dilysogens formed?

6 Replication of λ DNA can be detected by hybridization of labeled λ mRNA with all of the DNA isolated from an infected cell or from an induced lysogen. As replication proceeds and the number of copies of λ DNA increases, more λ mRNA can be hybridized. Interpret the following observations. (a) If an *xis*$^-$ lysogen is induced, the amount of hybridizable λ mRNA increases with time. (b) Also, by using *bio* mRNA and *gal* mRNA (both of which are easily obtainable) the amount of these mRNA's that can hybridize also increases with time. (c) If an *int*$^+$ lysogen is induced, the amount of hybridizable *bio* and *gal* mRNA (as well as λ mRNA) increases, although less of these hybridize than in part (b).

7 *E. coli* mutants, unable to synthesize thymine, were fed 5-bromouracil to make their DNA "heavy" and then were infected with the generalized transducing phage P1 and incubated in a medium containing (1) thymine, (2) 5-bromouracil, or (3) thymine plus ^{32}P until lysis occurred. Analysis of phage progeny by density-gradient centrifugation showed that transducing particles from the first two media have similar densities and that those from the third are not radioactive.

What do you know about the phage DNA and the origin of transduced segments?

8 λ*gal* can integrate into the left prophage attachment of a heteroimmune lysogen, not into the right prophage attachment site. Why not?

9 How might you make a λ lysogen unable to adsorb λ?

ANSWERS TO MORE CHALLENGING PROBLEMS

1 A nonsense mutation is in gene *cI*, *cII*, or *cIII* and B (not A) has a suppressor.

2 (a) The frequency with which a lysogen for a wild-type phage is spontaneously induced is relatively low, and the phage released form turbid plaques. However, if a mutation has occurred in the *cI* gene, a lysogen containing this mutant (which will form a clear plaque) will always produce phage. (b) The gene *cI* makes repressor. Genes *cII* and *cIII* are needed in the wild-type state for the establishment of lysogeny, but not for its maintenance. Thus, a *cII*$^-$ or *cIII*$^-$ mutation in a prophage does not result in escape from repression.

3 Insertion has occurred at the right prophage *att* site (that is, at *POB′*) to generate a dilysogen. The Ter system is active and cleaves at the two *cos* sites to produce a J^+P^+ recombinant. To prove this point, the prophage should be made A^+R^- and the infecting phage A^-R^+. All J^+P^+ phage will be A^+R^+.

4 A promoter-up mutation in *xis* would make *int* transcription constitutive.

5 Using the integrase of λ*b2* (λ*bio* has no *int* gene), an Int-promoted event has occurred yielding a dimeric circle with Δ*OB′* and *POP′* ($\Delta OP' \times POB' \rightarrow \Delta OB' + POP'$) and the dimer has integrated using the newly generated *POP′* site.

6 (a) Prophage DNA can replicate without excision. (b) The replication forks can leave the prophage and enter the bacterial DNA. (c) Excision must be fairly slow since often the replication forks leave the prophage before excision occurs.

7 The DNA in the particles replicated before infection. Also, either no bacterial DNA replicates after infection, or replicated DNA never gets into the particles.

8 Integration into *POB′* requires the Xis-mediated exchange (*BOP′* and *POB′*). In an infection, the Xis protein is not made to any great extent.

9 Phage-resistant mutants are obtained by plating 10^8 phage and bacteria and picking surviving colonies. With a λ lysogen all cells survive. However, if λ*imm434* were used, only adsorption-defective bacteria would form colonies.

CHAPTER 19

PLASMIDS

CHAPTER SUMMARY

Plasmids are extrachromosomal circular DNA molecules frequently found in bacteria and yeasts, and occasionally in higher eukaryotes. Their defining feature is their dispensibility; that is, only in rare or laboratory situations do they confer on their host any selective advantage. Furthermore, they have a replication origin and (probably always) carry one or more genes that regulate initiation of their replication. Except for initiation, they depend on host proteins for replication. The number of copies of a particular plasmid per cell is plasmid-specific. For some plasmids, the low-copy-number plasmids, the number of copies of a particular plasmid per cell is less than about 5, and sometimes 2, at the time of cell division. When a cell divides, these plasmids segregate in an orderly way into daughter cells. Other plasmids, the high-copy-number plasmids, are present in up to about 50 copies, and probably segregate at random. The DNA plasmids are invariably supercoiled. Their size range downward from a few percent of the size of the bacterial chromosome. A few RNA plasmids have been observed.

The three major classes of plasmids in *E. coli* are the (F) sex , R (drug-resistance), and Col (colicin-producing) plasmids. The F plasmid is self-transmissible and can be transferred from an F^+ cell to an F^- cell by bacterial conjugation. When integrated into the chromosome, forming an Hfr cell, F mediates transfer of the bacterial chromosome.

The plasmid-transfer process has four essential steps: (1) effective contact (formation of pairs), (2) mobilization (preparation for transfer), (3) DNA transfer, and (4) repliconation (formation of a functional plasmid in the recipient cell). These processes are controlled by plasmid-specific genes. A plasmid that can accomplish effective contact is said to be conjugative, and a plasmid having the second function is mobilizable. A plasmid that is both conjugative and mobilizable is self-transmissible. If a cell contains both a conjugative and a nonconjugative plasmid, the former can often provide the missing function for the latter, thereby allowing the nonconjugative plasmid to be transferred.

Effective contact usually requires the joining of a hairlike appendage (a pilus) on the donor cell to the recipient cell. Retraction of the pilus into the donor cell brings the cells into fairly firm contact. Mobilization involves nicking of the DNA and attachment of a transfer protein to the transfer origin. Looped rolling circle replication then occurs and a single strand is transferred to the recipient. Conversion to double-stranded DNA occurs in the recipient. Normal DNA replication of plasmids is by the θ mode and is not the same as transfer replication.

The number of copies of a particular plasmid per cell is determined by the concentration of a plasmid-encoded inhibitor of initiation of DNA synthesis. The inhibitor is usually active against all members of a given plasmid class. Thus, two members of the same class cannot usually be carried in a single cell; this phenomenon is called incompatibility. Plasmids are often classified according to incompatibility classes. The inhibitory mechanism is best understood for ColE1 DNA, for which replication has been carried out *in vitro*. Replication is by the θ mode. Initiation differs somewhat from that observed in phage systems and begins by synthesis of a long RNA molecule (RNA II) 651 bases upstream from the replication origin. For most of its length the RNA behaves like a typical transcript and is not base-paired with the DNA. Secondary structure within the free part of the RNA provides a signal for base-pairing to begin at a particular point. After the paired region is well past the replication origin RNase H makes a single cut at the origin, producing a 3'-OH group, which primes DNA synthesis. A regulatory RNA molecule, RNA I, is also synthesized from a point 110 bp downstream from the start of RNA II. It is complementary to RNA II in the region that signals base-pairing to begin and hence can inhibit initiation. A protein, Rom, mediates the pairing of RNA I and RNA II. How Rom works is unknown.

Inhibition of protein synthesis in a bacterium by an antibiotic prevents initiation of synthesis of the chromosome, presumably because of the need to synthesize unstable initiation proteins. Replication of the high-copy-number plasmids is not inhibited for two reasons: (1) their replication does not require any of the unstable bacterial products and (2) the plasmid inhibitor of initiation is not made in sufficient concentration to prevent initiation. The result is continued synthesis of the plasmid until several thousand copies are present in the cell. This phenomenon, which is called plasmid amplification, is useful in the preparation of large amounts of plasmid DNA.

F contains nucleotide sequences (insertion elements, described in Chapter 21) that are also present in the bacterial chromosome. Homologous DNA recombination occasionally occurs between a phage and a bacterial sequence, resulting in insertion of the plasmid into the chromosome and formation of an Hfr cell. Since F contains a transfer origin plus all of the elements needed for DNA transfer, the chromosome is, in effect, a giant self-transmissible plasmid and can be transferred to an F^- cell. Varied location of insertion elements in the chromosome account for the different positions of transfer origins in different Hfr strains.

The plasmid Ti of *Agrobacterium tumefaciens* is responsible for the formation of certain tumors in plants. Ti can be manipulated by genetic engineering techniques and has been used to transfer genes into plants. This technique provides a valuable means for developing new varieties of plants.

Eukaryotic plasmids are rare but have been observed in yeast and other fungi. The 2μm plasmid of yeast has two copies of a segment in an inverted-repeat array. Intraplasmid recombination between these segments causes an inversion. Recombination can occur a second time and restore the original sequence. Inversion and restoration occur repeatedly and presumably at random.

BOLD

amplification, colicin, conduction, conjugative plasmid, donation, donor conjugal DNA synthesis, effective contact, F' plasmid, F pili, incompatibility groups, incompatible, killer particle, lacuna, low-copy-number, mobilizable plasmid, mobilization, r determinant, R pili, recipient conjugal DNA synthesis, relaxation complex, relaxed, repliconation, RTF (resistance transfer factor), self- transmissible plasmid, sex pilus, stringent, transfer origin

ADDITIONAL INFORMATION

Various intercalating agents, particularly acridines (e.g., proflavin, acridine orange) can inhibit DNA replication in bacteria. For unknown reasons plasmids are inhibited by the acridines at concentrations that are too low to inhibit chromosomal DNA replication. Such inhibition can lead to loss of the plasmid and is called acridine curing. It is detected most easily when the plasmid contains host genes. If a culture of a bacterium whose genotype is F'lac^+/lac^- (that is, the chromosome is lac^- while F carries the lac^+ allele) is grown in media containing acridine orange, cells continue to grow and divide. However, after one generation the number of Lac$^+$ cells remains constant and the number of Lac$^-$ cells increases. The explanation is the following. At the time of adding acridine orange, most cells contain two copies of F'*lac*. After one generation each cell contains only one F'*lac* because plasmid replication is inhibited. Thus, in the next cell division only one plasmid is available for the two daughter cells; as a consequence, plasmid-free bacteria are produced.

DRILL QUESTIONS

1 What is meant by a plasmid?

2 What kind and what form of nucleic acid comprises plasmids?

3 What is the size range in base pairs of plasmid DNA?

4 Is there any relation between plasmid size and copy number?

5 What is a relaxation complex?

6 When DNA is isolated from cells, nicks are often produced by various nucleases. How does the nicking produced by a relaxation complex differ from the nicking of chromosomal DNA?

7 What mode of replication is used by plasmids for their reproduction?

8 Are there any circumstances in which plasmids undergo rolling circle replication?

9 What early enzymatic step is needed in transfer replication, but not in normal replication?

10 During transfer, does replication occur in the donor, the recipient, or both?

11 What early enzymatic step is required in transfer replication but is not required for typical θ replication?

12 Are sex pili present on the donor or recipient cell?

13 What element encodes the genes for pili synthesis?

14 What two roles are played by pili in bacterial conjugation?

15 What are the *tra* genes of F?

16 (a) Define the terms conjugative, mobilizable, and self-transmissible. (b) Which of these terms describe F? (c) Which of these terms describes ColE1?

17 If a plasmid is mobilizable but nonconjugative, what function does it lack?

18 In a hybrid plasmid carrying two replication origins, one from a high-copy-number plasmid and the other from a low-copy-number plasmid, which origin will be used and what will the copy number be?

19 If two plasmids cannot be maintained in a single cell, what property is common to the plasmids?

20 How does RNase H function in replication of ColE1? What property does RNase H have, as opposed to a typical RNase, which enables it to perform this function?

21 What feature of F, not present in ColE1, enables F, but not ColE1, to integrate

into the host chromosome?

22 Will an Hfr strain have sex pili?

23 If ColE1 could be altered to contain insertion sequences homologous to sequences in the chromosome, such that it could integrate in the chromosome, would Hfr-like cells arise?

24 What are the two components of R factors and which carries the genes for replication and transfer?

ANSWERS TO DRILL QUESTIONS

1 A circular DNA—and in one case, RNA—molecule (usually much smaller than the host chromosome) that replicates within a host cell doing the cell no particular harm and, in most environmental conditions, no good either.

2 Supercoiled double-stranded DNA, except for the RNA killer particle of yeast.

3 The molecular weights range from a few million up to about 100 million, so the number of base pairs ranges from about 1500 to about 150,000.

4 Not really. The largest plasmid have a low copy number but the small plasmids can exist in any number.

5 A plasmid to which plasmid-specific proteins, one of which has nuclease activity, are bound.

6 Nicking of a relaxation complex occurs in one strand and at a particular site.

7 θ replication.

8 Yes, during transfer.

9 A nick producing a 3'-OH terminus.

10 In both. Transfer replication occurs in the donor and conversion of the transferred single strand to double-stranded DNA occurs in the recipient.

11 A nick at the transfer origin to generate a 3'-OH group.

12 The donor.

13 The genes reside in the plasmid. Thus, F pili are made by pili genes in F, and R pili are made by genes in an R factor.

14 (1) They are involved in recognition of donor and recipient; and (2) the donor pilus attaches to the recipient cell, retracts and brings the two cells in close contact.

15 The 19 genes required for transfer, for example, pili genes, and replication genes.

16 (a) A conjugative plasmid is able to form a donor-recipient pair. A mobilizable plasmid carries genes needed for DNA transfer. A self-transmissible plasmid is both conjugative and mobilizable. (b) All three. (c) Conjugative.

17 It cannot mediate effective contact (formation of donor-recipient pairs).

18 The origin from the high-copy-number plasmid. Hence, the copy number will be high.

19 They must have the same, or at least similar, inhibitors of replication initiation.

20 RNase H nicks the RNA primer at the replication origin, producing a 3'-OH group. RNase H can nick RNA in a DNA-RNA hybrid; such RNA is resistant to most RNases.

21 F contains DNA sequences that are homologous to sequences in the chromosome. Recombination between these sequences gives rise to insertion.

22 Yes, since it contains F, which has the pili genes.

23 No. ColE1 is nonconjugative so the cell could not transfer its DNA unless F were also present.

24 RTF, which carries the replication and transfer genes, and the r determinant, which carries the genes for antibiotic resistance.

PROBLEMS

1 An F'*lac* donor cell is mated with an F^- cell carrying a *dnaG*(Ts) mutation at a nonpermissive temperature. Could β-galactosidase be made by the recipient?

2 A culture of $F'gal^+/gal^-$ cells is grown for many generations in medium containing acridine orange. At various times cells are taken and plated on a color-indicator plate that distinguishes Gal$^+$ from Gal$^-$ cells. After a few generations, the fraction of Gal$^+$ cells remains constant at about 0.1% of the

population. If Gal$^+$ colonies are taken and recultured in medium containing acridine orange, Gal$^-$ cells do not appear. If the original bacterial strain was RecA$^-$, the fraction of cells that is Gal$^+$ continues to drop until no Gal$^+$ cells can be found. Explain this phenomenon.

3 Could you have a plasmid with no genes whatsoever?

4 How do the sizes of plasmids compare with the size of phage DNA molecules?

5 Lac$^+$ and Lac$^-$ cells form purple and white colonies respectively on EMB lactose agar. A culture consisting of cells whose genotype is F'*lac*$^+$/*lac*$^-$ is mutagenized and plated on EMB-lactose agar at 30 C. Several purple colonies whose color is a little less intense than the others are studied further. These give white colonies at 42 C and light purple colonies at 30 C. Bacteria obtained from white (42 C) colonies remain white when replated at 30 C. What type of mutation has been acquired?

6 (a) An *E. coli* cell containing F can transfer the chromosome at a frequency of about one transfer event per cell. Is this donation or conduction, and does it occur in a RecA$^-$ cell? (b) Can ColE1 cause chromosome transfer?

7 If, in a particular cell type, rifampicin were to inhibit DNA transfer, what would you conclude about the transfer mechanism?

8 What are the roles of DNA synthesis in the donor and the recipient?

9 What procedure can be used to transfer a nontransmissible plasmid from one strain to another?

10 Although the concentration of enzymes that are synthesized is tightly regulated, it is often found that more enzyme is made per cell if an operon is located on an F' sex plasmid than if it is located on a bacterial chromosome. Explain.

11 Would you expect an Hfr strain of *E. coli* to have sex pili?

12 A Lac$^-$ bacterial strain has a *dnaA*(Ts) mutation, which prevents colony formation at 42 C. An F'*lac*$^+$ plasmid is introduced into the strain by conjugation. The culture is grown for many generations at 30 C, and then 10^6 cells are plated at 42 C. A few colonies arise and these are capable of growth at both 30 C and 42 C. (a) Are these colonies Lac$^+$ or Lac$^-$? (b) Do the cells still carry the *dnaA*(Ts) mutation? (c) What feature of the cell has changed that enables them to grow at 42 C? (d) Can the cells grow at 4 C in the presence of acridine orange?

13 An F'(Ts)*lac*$^+$ plasmid has a temperature-sensitive mutation in its replication system. (a) What is the phenotype of an F'(Ts)*lac*$^+$/*lac*$^-$ cell at 42 C? (b) An F'(Ts)*lac*$^+$/*lac*$^-$*gal*$^+$ strain is grown for many generations and then plated at 42 C. Some Lac$^+$ colonies form at 42 C. How have these formed? (c) Some of the Lac$^+$ colonies in (b) are Gal$^-$. How have these formed?

14 A strain carrying F'*gal*$^+$, which forms red colonies on MacConkey-galactose agar (Gal$^-$ colonies are white), is mutagenized and plated. A few colonies are found that are slightly smaller and more intensely red. Further study shows that they have 10 copies of F'*gal*$^+$ per cell rather than the usual number. What types of mutations have occurred?

15 Plasmids contained in Rec$^+$ cells frequently dimerize. A dimer can then be isolated and transferred by the $CaCl_2$ technique to a Rec$^-$ cell in which it will be stable and not revert to the monomer. If a monomer plasmid has a copy number of 10, what will the copy number of the dimer be?

16 On EMB-lactose agar Lac$^+$ and Lac$^-$ colonies are purple and white respectively. If several thousand F'*lac*$^+$/*lac*$^-$ cells are plated, a few sectored colonies appear. This may have a purple half and a white half or a wedge of white in a predominately purple colony. Sectored colonies that are predominately white are not found. Explain the cause of sectoring.

17 Complete transfer of the chromosome in Hfr transfer requires about 100 minutes. (a) At what time(s) is (are) F DNA transferred? (b) A Str-s Hfr that transfers the genes in alphabetical order, *a b ... y z*, is mixed with a *F$^-$z$^-$str-r* cell. The mixture is agitated violently 15 minutes after mixing to break apart conjugating cells and then plated on agar lacking Z and containing streptomycin. About one cell in 10^7 yields a Z$^+$Str-r colony. What is the genotype of such a colony?

ANSWERS TO PROBLEMS

1 No. A single strand of F'*lac* would be transferred by normal looped rolling circle replication. However, without functional primase, the transferred circle would not be converted to double-stranded DNA. Hence the *lac* gene would not be transcribed and enzyme would not be made.

2 Acridine causes inhibition of replication of the plasmid, leaving cells that are plasmid-free, i.e., Gal$^-$. However, some genetic recombination can occur by which the plasmid becomes inserted into the chromosome. Such a plasmid is then replicating under control of the bacterial replication system, so the cell remains Gal$^+$. If the cell is RecA$^-$, recombination and consequent plasmid insertion cannot occur.

3 Yes. The simplest plasmid would be a tiny DNA fragment carrying no infor-

mation other than a copy of the replication origin of the host cell.

4 They are comparable.

5 The plasmid has a temperature-sensitive mutation in some function required for DNA replication.

6 (a) Conduction; no. (b) No, it is not self-transmissible and the chromosome is neither mobilizable nor conjugative.

7 Since an RNA primer is not needed to initiate rolling circle replication, a reasonable guess would be that transcription of an inducible gene is needed for transfer.

8 Synthesis in the donor provides the single strand that is copied. Synthesis in the recipient converts the transferred strand to double-stranded DNA.

9 $CaCl_2$ transformation using purified plasmid DNA.

10 There are frequently two copies of an F' plasmid per cell.

11 Of course, because the expression of F genes is the same in the integrated and nonintegrated states.

12 (a) Lac$^+$. (b) Yes. (c) F'*lac* has integrated into the chromosome. The chromosome is now replicating from the origin of F and hence does not need the DnaA protein. (d) No.

13 (a) Lac$^-$. (b) F'(Ts)*lac*$^+$ has integrated into the bacterial chromosome. (c) Integration has occurred within a *gal* gene.

14 Either the replication repressor or the repressor binding site has been altered. This means that a higher repressor concentration is needed to shut off replication.

15 5.

16 Early in the development of the colony a cell has divided to yield a daughter cell lacking F'*lac*.

17 (a) Since the transfer origin is in F and F DNA is transferred linearly, some F DNA is transferred first (just after $t = 0$) and the remainder is last (at $t = 100$). (b) Some of the cells in the Hfr population must have formed F' plasmids for transfer of a "late" gene to have occurred early. Thus, the genotype is F'z^+/z^- *str-r*.

MORE CHALLENGING PROBLEMS

1 In general an Hfr cell and an F^+ cell cannot mate because pair formation does not occur. Growth of a male cell (Hfr or F^+) in anaerobic conditions for a period of time converts it to a "phenotypic female"—one that lacks pili and can pair with a normally grown male. It is possible by the detection of genetic recombinants to observe the transfer of DNA from an Hfr to an F^+ phenotypic female. However, no genetic test has demonstrated transfer of the F plasmid to an Hfr phenotypic female. What is the cause of this asymmetry?

2 Replication of unintegrated F, but not integrated F, is inhibited by exposing *E. coli* to acridine orange. Make use of this finding (a) to obtain F^- from F^+ cells, and (b) to identify colonies as F^+, Hfr, or F^-.

3 With certain combinations of donor and recipient cells plasmid transfer is very inefficient. For example, if two *E. coli* strain K cells (one F'*lac*$^+$ and one F^-*lac*$^-$) are mated for 15 minutes, the number of F^-*lac*$^-$ cells converted to Lac$^+$ is roughly equal to the number of F'/*lac*$^+$ cells. However, if the F' is contained in strain K and the F^- cell is strain B, the number is about 10^5- fold smaller. These strain F'*lac*$^+$ B cells can then be used in subsequent matings. When this is done, it is found that in a B x B mating, the transferred plasmid is detected in both F^- cells, whereas in a B x K mating, the frequency is again quite low. Explain this phenomenon.

ANSWERS TO MORE CHALLENGING PROBLEMS

1 Whereas F may in fact be transferred to an Hfr phenotypic female, incompatibility (namely, the F replication repressor) prevents F from being maintained in a colony that might develop from an Hfr cell that had received F. Incompatability is not a factor in detecting Hfr DNA because by recombination the transferred DNA becomes part of the chromosome of the F^+ cell.

2 (a) Grow the F^+ culture on agar containing acridine orange. Test colonies either for inability to transfer markers, ability to be a recipient or resistance to male-specific phages. (b) F^- cells can act as a recipient, which immediately distinguishes them from donors. Growth of F^+ but not Hfr cells in media containing acridine orange leads to the formation of cells which act as recipients.

3 This is the host modification-restriction phenomenon first observed with phages. In a K x B or a B x K mating the transferred DNA is enzymatically destroyed in the recipient. In the few recipient cells in which restriction has not occurred the plasmid DNA gets the new host modification.

CHAPTER 20

GENETIC RECOMBINATION BETWEEN HOMOLOGOUS DNA SEQUENCES

CHAPTER SUMMARY

There are two basic types of recombination: homologous, in which participating base sequences are the same or nearly the same, and nonhomologous, in which the sequences are different. Homologous recombination requires a protein of the type called RecA in *E. coli* for catalyzing strand invasion and pairing of complementary sequences. Nonhomologous recombination utilizes enzymes that are either responsible for joining two particular sequences (site-specific recombination) or for joining a specific sequences with an arbitrarily selected nonhomologous sequence (transposition). In illegitimate recombination distinct sequences are broken and rejoined in a completely unknown way. Breakage, strand exchange, and reunion of DNA fragments occur in all modes of recombination. Formation of heteroduplex regions and DNA replication in which gaps are filled are characteristic of homologous exchange.

The RecA protein is responsible for the primary association of single- and double-stranded DNA. The reaction begins with a presynaptic coating of single-stranded DNA with RecA protein. In the presence of ATP the RecA nucleoprotein associates with double-stranded DNA; the association occurs without regard for sequence homology. Nonhomologous single- and double-stranded regions drift back and forth with respect to one another until complementary sequences are aligned. A presumed unwinding activity of RecA causes assimilation of the single-stranded DNA into the double helix, with associated base-pairing; the mechanism of strand assimilation is not known. Another enzyme, the RecBC enzyme, is also essential for homologous recombination in *E. coli*, but its function is unknown.

In all modes of homologous recombination strand breakage is an early event that leads to formation of a heteroduplex region, consisting of one strand from each of two participating DNA molecules. The heteroduplex region, which results by strand assimilation, need not contain the initially paired segments, since branch migration enables the region of pairing to migrate. Strand exchange occurs some time after pairing. Following strand exchange the heteroduplex region may contain a base-pair mismatch if each of the two strands contains one base of a mutant and of a wild-type base pair. The mismatch may be repaired, leading to loss of one of the genetic markers, or the region may replicate without mismatch repair.

Bacterial recombination begins with uptake of a double-strand DNA fragment. Accompanying uptake is degradation of one of the single strands, so only a single strand enters the cell. This strand often undergoes RecA-mediated pairing, according to the mechanism described above, resulting in incorporation of the donor single-strand into the double-stranded chromosome of the recipient cell. Strand assimilation, branch migration, trimming of free ends, and ligation completes the incorporation process. If mismatch repair occurs, the donor marker may be either lost or may completely replace the marker of the recipient. If mismatch repair does not occur, replication of the donor cell yields one wild-type and one mutant cell.

Meiotic recombination has been investigated by study of the ascus-forming fungi. Unusual asci occasionally arise as a result of certain events occurring during meiosis. Crossing over is a reciprocal event, so two molecules with reversed heteroduplex regions form. These aberrant asci, in which genetic markers do not segregate in 4:4 Mendelian fashion, result from the fact that each of the two heteroduplex regions contains different alleles. Since most alleles differ by a single base pair, a mismatched pair of bases is present in each heteroduplex region. These mismatches can be corrected by a repair system. In fungi with ordered asci, if mismatch repair does not occur, asci are formed with two nonidentical pairs of spores. The alleles are still present in a 4:4 ratio, but because of the aberrant order, the asci are called aberrant 4:4 asci or are said to show aberrant 4:4 segregation. When both mismatched pairs are repaired in the same direction (mutant to wild-type or wild-type to mutant), all spore pairs are identical, but the alleles are present in a 6:2 or 2:6 ratio; this phenomenon is called gene conversion or 6:2 segregation. If mismatch repair occurs in only one heteroduplex, an ascus forms in which the members of only one pair of spores have different genotypes; such an ascus is a postmeiotic-segregation ascus and exhibits 5:3 segregation.

Most models of recombination between double-stranded DNA molecules include a crossed-strand intermediate called a Holliday junction. In the junction two double-stranded molecules are joined by nearby homologous single strands. The crossed-strand region migrates by branch migration until pairs of cuts (the resolution step) are made that separate the double-stranded molecules. The positions of the paired cuts determine whether the separated molecules are recombinant or nonrecombinant. Several models suggest means for producing a Holliday junction. An important model is the asymmetric strand-transfer model. In this model a nick is made in one molecule and displacement replication, by an enzyme like polymerase I, produces a single-stranded appendage. By means of RecA-mediated synapsis this appendage invades the unnicked strand, forming a D loop. The D loop is then cleaved, resulting in strand assimilation. An isomerization step creates the crossed strands. Branch migration and resolution determine whether a nonrecombinant or a recombinant molecule result.

Recombination between circular DNA molecules generates a figure-8 molecule with a Holliday junction at the joint. The resolution step separates the individual circles. Cuts in certain positions can also produce a rolling circle, but the significance of this possibility is not known.

BOLD TERMS

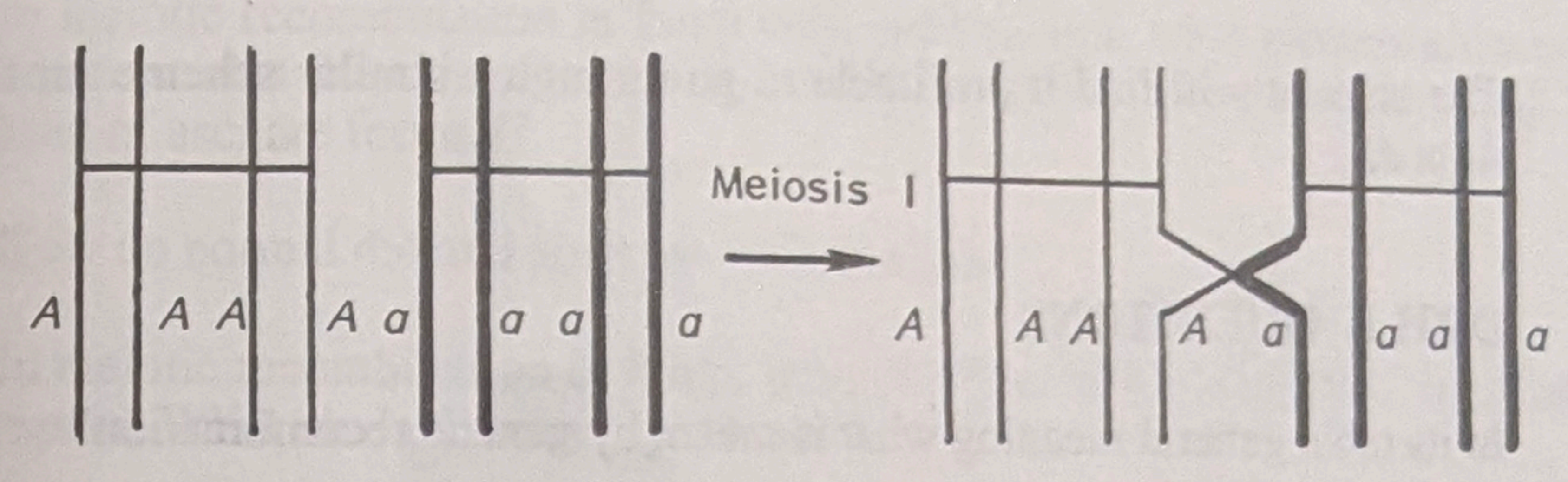
aberrant 4:4, ascus, assimilation, asymmetric strand-transfer model, co-conversion, conjugation, converted, crossing over, double-strand-break-repair model, 4:4 segregation, gene conversion, Holliday junction, homologous, homologous alignment, homologous recombination, invasion, postmeiotic segregation, resolution, strand invasion

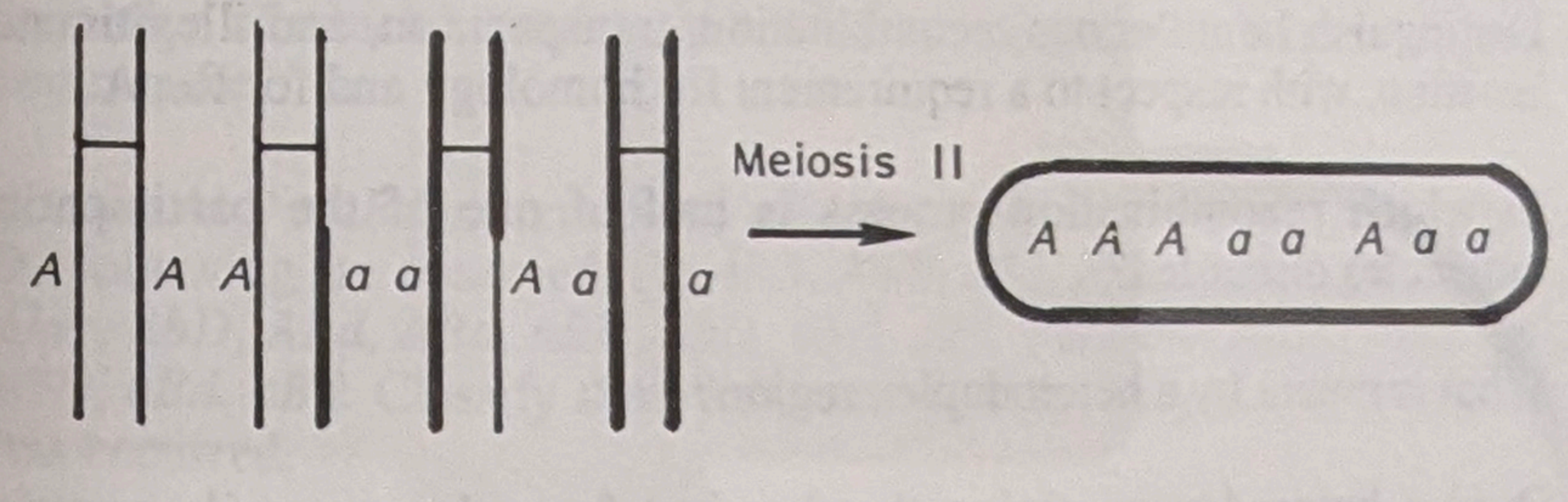
ADDITIONAL INFORMATION

In the early 1950s, shortly after the elucidation of the structure of DNA, an experimental approach was taken to understand the molecular mechanism of genetic exchange. It did not take long to recognize that there are many modes of exchange, not only in different organisms but within the same organism. The multiplicity of mechanisms is most evident in *E. coli*, in which one can see transformation, conjugation, transduction, and so forth. However, when one looks closely at *E. coli*, it is found that the multiplicity is even greater than the number of obviously different phenomena. For example, in conjugation it is possible to eliminate about 99% of the recombination by a mutation in a critical gene called *recB*. However, some recombination remains, and to eliminate that completely requires mutations in additional genes: *recE*, *recF*, and so forth. It is clear that at least three independent mechanisms are active in the overall processes of conjugation and transduction, all of which depend on an active RecA protein. When phage λ infects *E. coli*, phage-phage recombination occurs; this exchange does not require the RecA protein, but utilizes two phage gene products called Red and Beta. Phage T4 has a recombination system of its own also. Furthermore, careful analysis of the recombination that occurs in conjugation and with λ and T4 and indicates that these are quite different processes. For some time, it was assumed that although there are many recombination systems, which are more or less important in different organisms, the major recombination system in related microorganisms might be the same and that this system might be the one used in meiotic recombination. Without going into detail it is clear that this is not the case. If there is a relation between meiotic recombination and any system in *E. coli*, it is at best with one of the minor systems. Similarly, the major system in λ resembles a minor one in *E. coli*. Nonetheless there do seem to be common features in all systems: the enzymes

used, the existence of Holliday junctions, and so forth. The general message to be learned is that one can not always utilize experimental results from one system to understand another system.

Usually students find the subjects of postmeiotic segregation and gene conversion to be somewhat baffling. Actually, they are quite straightforward and easily understood if one takes the time to write out all of the possibilities in a particular ascus. It should be noted though that simple crossing over is not sufficient to generate these "unusual" asci, for their production requires that the position of the heteroduplex region is such that dissimilar alleles are included. We will now present the simplest case, a cross between *A* and *a* in which the heteroduplex includes *A* and *a*. With only one marker recombination is not evident by the presence of new genotypes; however, the formation of unusual asci will make it clear that crossing over has occurred. All possibilities will be written out. Figure 20-12 in the text shows that in the absence of any crossing over the spore order in an ordered ascus is *AAAAaaaa*, that is, 4:4. The result of a crossover that generates *A*/*a* heteroduplex regions is shown in the figure below:





An exchange occurs as shown, so that after meiosis I, chromosomes 2 and 3 (arbitrarily numbered) contain the heteroduplex regions. Note that in the absence of any mismatch the spore order is *AA Aa aA aa;* the composition is 4 *A*:4 *a*, with two unmatched spore pairs (that is, *Aa* and *aA*), so this is an aberrant 4:4 ascus. On the following page we list the asci that would result with the mismatch repair events indicated.

Correction	*Spore order*				*Ascus type*
One *A* to *a* in chromosome 2	*AA*	*aa*	*aA*	*aa*	3:5
One *A* to *a* in chromosome 3	*AA*	*Aa*	*aa*	*aa*	3:5
One *a* to *A* in chromosome 2	*AA*	*AA*	*aA*	*aa*	5:3
One *a* to *A* in chromosome 3	*AA*	*Aa*	*AA*	*aa*	5:3
a to *A* in both 2 and 3	*AA*	*AA*	*AA*	*aa*	6:2
A to *a* in both 2 and 3	*AA*	*aa*	*aa*	*aa*	2:6
a to *A* in 2, and *A* to *a* in 3	*AA*	*AA*	*aa*	*aa*	4:4 (normal)
A to *a* in 2, and *a* to in 3	*AA*	*aa*	*AA*	*aa*	4:4 (normal)

The student will find it profitable to go through a similar scheme for the cross *Ab* x *aB*.

DRILL QUESTIONS

1 In its most general meaning what is meant by genetic recombination?

2 Are most forms of exchange homologous or nonhomologous?

3 Distinguish homologous recombination, transposition, and illegitimate recombination, with respect to a requirement for homology and for RecA.

4 In which recombination process is half of one of the participating DNA molecules discarded?

5 What is meant by a heteroduplex region?

6 Does a heteroduplex region that has just formed necessarily contain a mismatched base pair?

7 Does branch migration occur in a particular direction, for example, 5'-to-3' or 3'-to-5'?

8 If mismatch repair of a heteroduplex occurs, how many types of daughter DNA molecules result?

9 During ordinary DNA replication mismatch repair following misincorporation usually yields the parental base sequence. In mismatch repair of a heteroduplex, is wild-type preferred over the mutant base pair?

10 In mismatch repair of a heteroduplex is one member of the pair ever preferentially used as a template?

11 List two biochemical activities of the RecA protein.

12 What steps in the RecA-mediated synapsis process could occur between a double-stranded circle and a completely nonhomologous single strand?

13 In RecA-mediated synapsis has base-pairing occurred in the stage of homologous alignment?

14 Name two properties of RecA that are important for recombination and state the stage at which each is important.

15 In meiotic recombination in fungi with ordered asci, what phenomenon clearly demonstrates the existence of a heteroduplex with a base-pair mismatch? What kind of asci are formed?

16 How do normal 4:4 and aberrant 4:4 asci differ?

17 In meiotic recombination in fungi, what phenomenon is explained by mismatch repair? What kind of asci are produced?

18 A cross is carried out in *Neurospora* between a wild-type (*A*) strain and a mutant (*a*) strain. Some of the asci are *AAAAaaaa* and others are *AAaaAAaa*. In the formation of which ascus has there been crossing over between the gene and the centromere?

19 A cross *AbD* x *aBd* is carried out with *Neurospora*. Among several types of asci the following are observed: (1) *AbD, AbD, Abd, Abd, aBD, aBD, aBd, aBd*; (2) *AbD, AbD, ABd, ABd, aBD, abD, aBd, aBd*; (3) *AbD, AbD, ABd, ABd, aBD, aBD, aBd, aBd*. Classify these asci, and identify the pairs in which conversion has occurred.

20 How many mismatch corrections occur in a 6:2 asci, and what can be said about the direction of the correction?

21 In what way is the replication that occurs in the asymmetric strand-transfer model different from normal DNA replication?

22 In the asymmetric strand-transfer model what enzyme is responsible for strand

assimilation, and could an enzyme like DNA polymerase III suffice?

23 All types of recombination processes (except for Int-promoted recombination) are terminated by the same enzymatic event. What is that event? What happens that is different in Int-promoted recombination?

24 Which types of recombination necessarily require two exchanges?

ANSWERS TO DRILL QUESTIONS

1 A rearrangement of the base sequence of DNA.

2 Homologous.

3 *RecA* and homology are needed only for homologous recombination.

4 Bacterial transformation.

5 A double-stranded segment formed from single-stranded regions from different DNA molecules.

6 Not if it forms between strictly homologous regions. In drawing models of recombination one usually draws the strand exchanges between genetic markers, so the heteroduplex region contains the two markers. However, exchange occurs at random, and more often than not it will occur in regions of complete homology. Such exchanges will of course be undetectable and are not particularly interesting from a genetic point of view, but they nonetheless occur. In carrying out physical experiments to study exchange, these nonrecombinational exchanges must of course be considered.

7 No. Migration occurs in both directions with equal probability.

8 One.

9 No.

10 Yes, but it may be either the wild-type or the mutant. Presumably nearby sequences affect the direction of repair.

11 DNA-binding, proteolysis.

12 Presynaptic binding and conjunction. Drifting could occur but homologous alignment would never be achieved.

13 No. Base-pairing occurs in the next stage.

14 DNA-binding is necessary for conjunction. Helicase activity is necessary for strand assimilation.

15 Postmeiotic segregation. Aberrant 4:4 asci.

16 In a normal ascus both members of each spore pair are identical; in an aberrant 4:4 ascus two spore pairs have different members, though the overall ratio of alleles for each gene is 4:4.

17 Gene conversion; 5:3 and 6:2 asci.

18 The one with the gene order *AAaaAAaa*.

19 (1) is a normal ascus; no conversion events. (2) is a 5:3 ascus; in the second pair of spores a *b*-to-*B* conversion has occurred; (3) is a 6:2 ascus; in the second and third pairs of spores a *b*-to-*B* conversion has occurred.

20 There are two corrections and both are in the same direction, that is, both x to X, or both X to x.

21 No initiation protein or replication origin is needed.

22 An enzyme of the pol I type. No, because the enzyme must be able to act at a nick and must have displacement activity, neither of which is possessed by pol III.

23 Ligation by a DNA ligase. Integrase carries out the ligation itself.

24 Transduction and bacterial conjugation.

PROBLEMS

1 What are the differences between transduction, transposition, transformation, bacterial conjugation, site-specific recombination, and meiotic recombination, with respect to requirements for homology and for a RecA-type protein?

2 In a reciprocal exchange no DNA is lost and all products of the exchange are recovered in progeny. Which of the following are examples of reciprocal exchanges? Prophage integration, prophage excision, formation of recombinants in conjugation, transduction by phage P1, formation of an F' plasmid from an Hfr bacterium, formation of an Hfr bacterium from an F^+ cell, bacterial transformation, and meiotic crossing over.

3 Irradiation with ultraviolet light leads to DNA damage, which is repaired. An early stage in the repair process is the formation of a gap in the polynucleotide

strand at the site of the damage. Such irradiation is also known to affect recombination frequency. If the repair system is inactive, owing to mutation, irradiation has very little effect on recombination frequency. What do you think the effect is, an increase or decrease of recombination frequency? Explain your answer.

4 A Lac⁻ *Pneumococcus* is treated with DNA from a Lac⁺ strain and plated on a nonselective color-indicator medium in which Lac⁺ and Lac⁻ cells produce red and white colonies, respectively. What types of colonies will form on the medium?

5 The exchange mechanism needed in transduction and in Hfr x F^- crosses was not discussed in the text. Would you expect them to be like transformation or like each other?

6 In a Hfr *leu⁺str-s* x *F⁻leu⁻str-r* mating in which the *leu* gene is transferred early, there are roughly 40 *leu⁺str-r* recombinants per 100 Hfr cells. If the F^- cells are also *recA⁻*, the number of such recombinants is about 10^{-6}. This reduction is true for all early markers. However, if one studies markers transferred very late, for which the recombination frequency is generally much lower than 40 per 100, it is observed that a *recA⁻* mutation in the recipient cell does not alter the recombination frequency at all. Explain this lack of reduction, which is a general feature of the transfer of late markers to a RecA⁻ strain.

7 What observation in meiotic crossing over is one of the best arguments for the existence of heteroduplex regions in recombination intermediates?

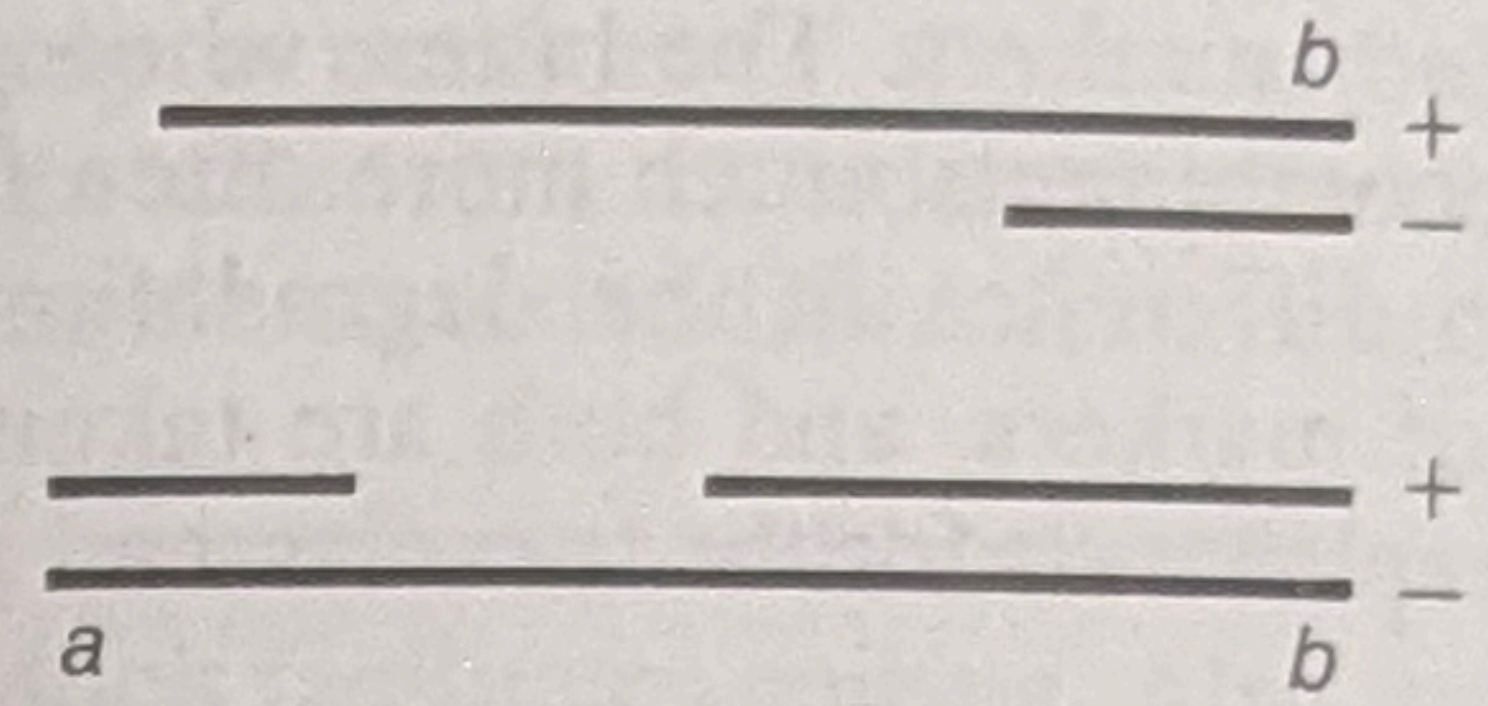
8 Haploid yeast are of two mating types, *a* and α, and these can fuse to form a diploid. Their genomes are identical except for the genes of the mating-type locus. A population of diploids can be grown on certain media and sporulation will occur. A type-*a* yeast with the genotype *D e f* (gene order *d e f*) is mated with a type-α *d E F* yeast and a diploid culture is obtained from a single zygote. The diploid culture is sporulated and asci are collected. Sixty-two asci contain 2 *D e f* spores and 2 *d e F* spores (these can be called type-I asci for convenience). Four asci contain 1 spore each of genotypes *D e F*, *D e f*, *d E F*, and *d E f* (type-II asci). Two asci contain 1 spore each of genotypes *D e F*, *D e F*, *d e F*, and *d E f* (type III). Explain how each class (I, II, and III) arose.

9 A fungal mutant is found in which aberrant 4:4 asci are observed but no 3:5, 5:3, 6:2, or 2:6 asci are ever found. What system is probably lacking in this mutant?

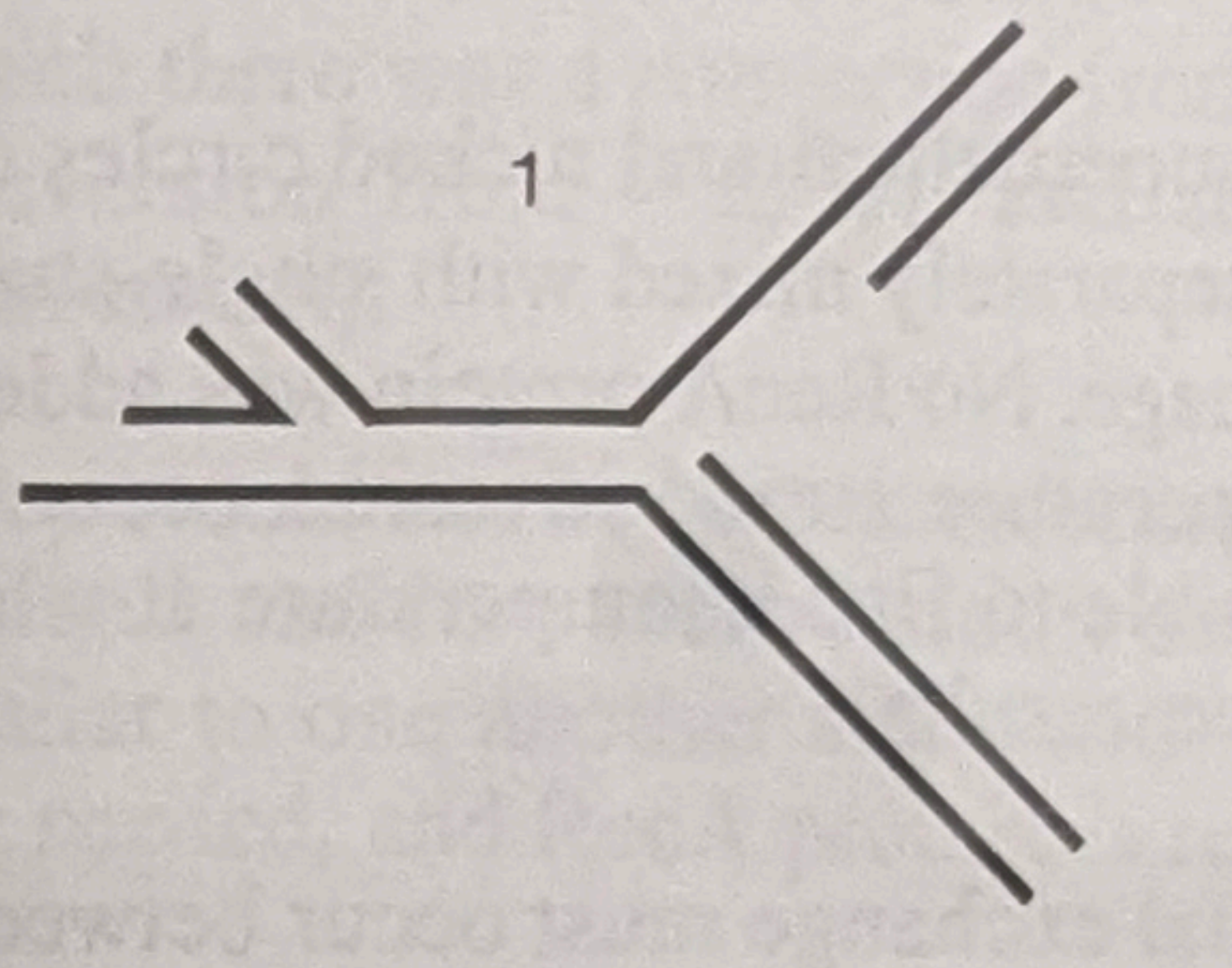
10 Consider a double-stranded circular DNA molecule and a single-stranded fragment, half of which is homologous to a sequence in the circle and half of which is totally different. The molecules are mixed, heated to a temperature at which the circle is partially denatured, and then cooled slowly. The sample is examined by electron microscopy and circles with single-stranded appendages of various length are seen. The next day fresh samples are prepared for microscopy. No circles with appendages are seen. Instead, all circles appear to be normal double-stranded circles. Free single strands are seen also. Explain these observations.

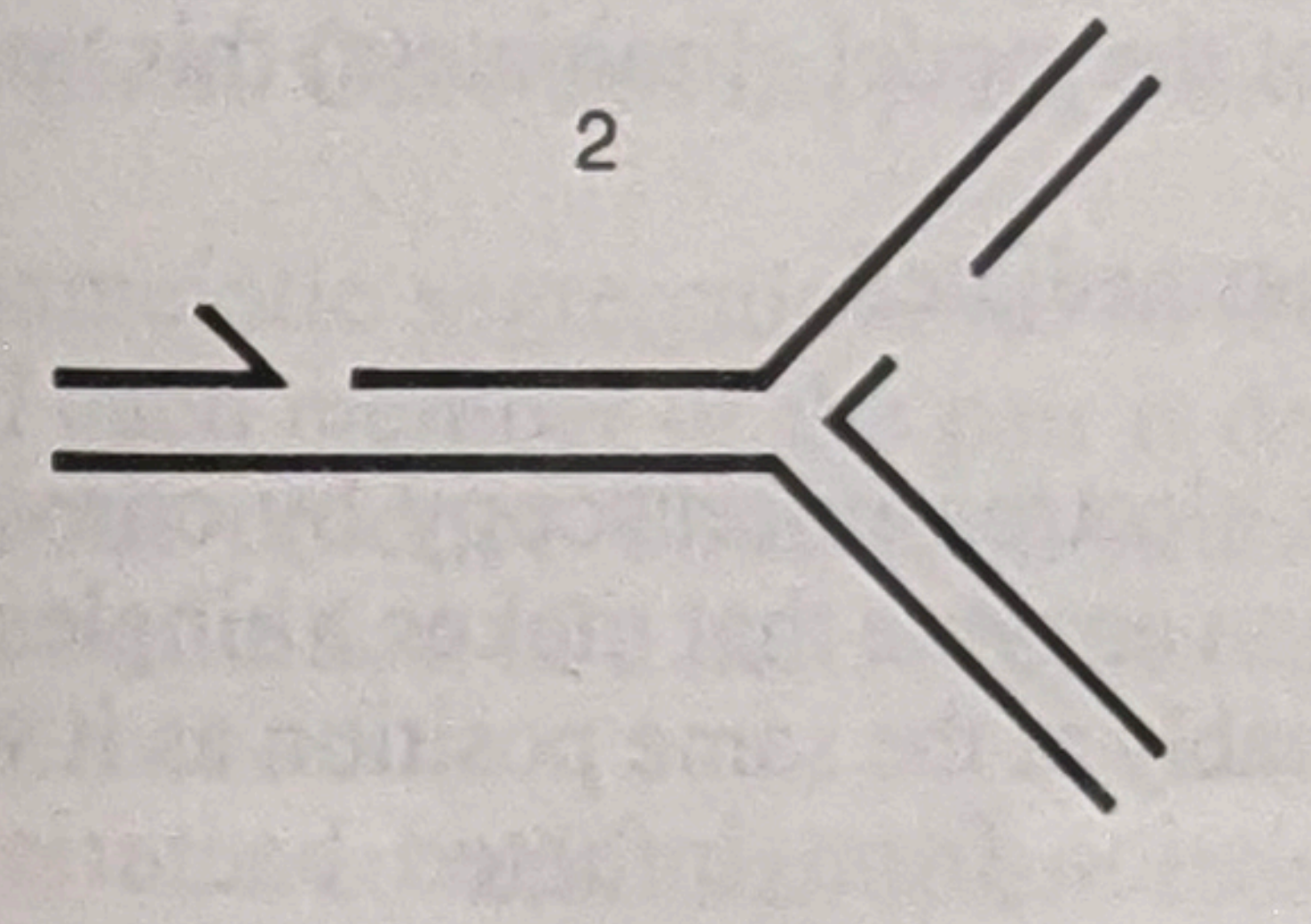
11 Both theoretically and experimentally recombination frequency can never exceed 50%. How is this explained by either the Holliday or asymmetric strand-transfer models?

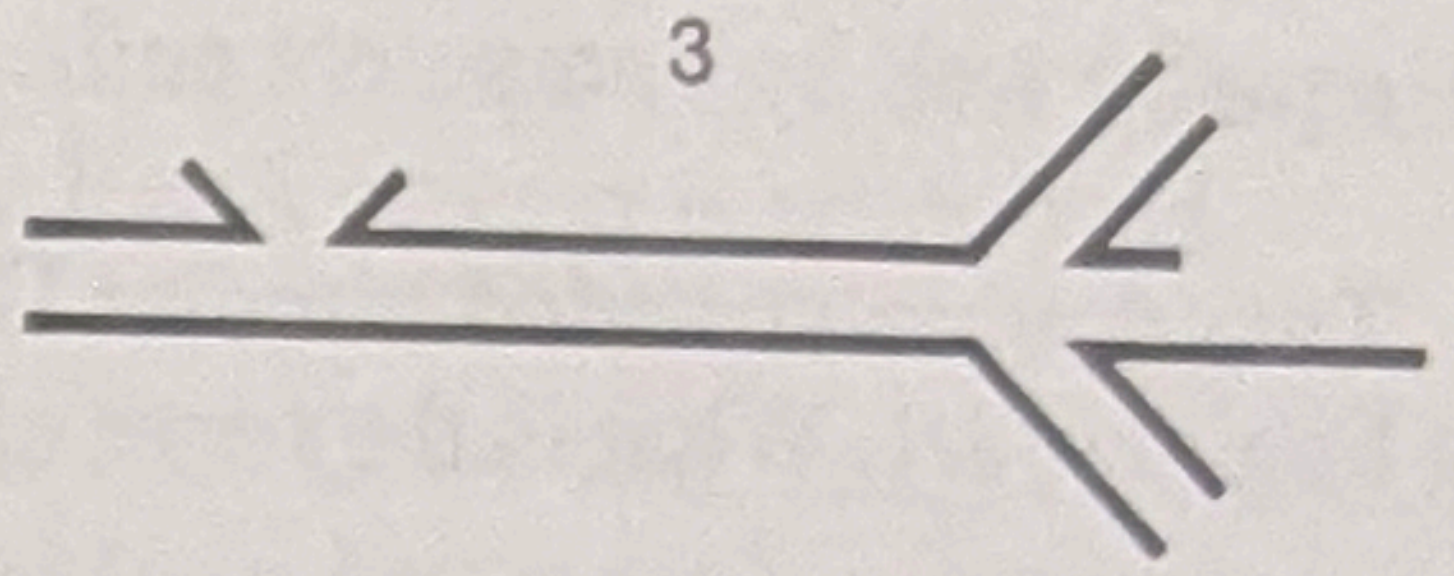
12 Two homologous DNA structures are shown below

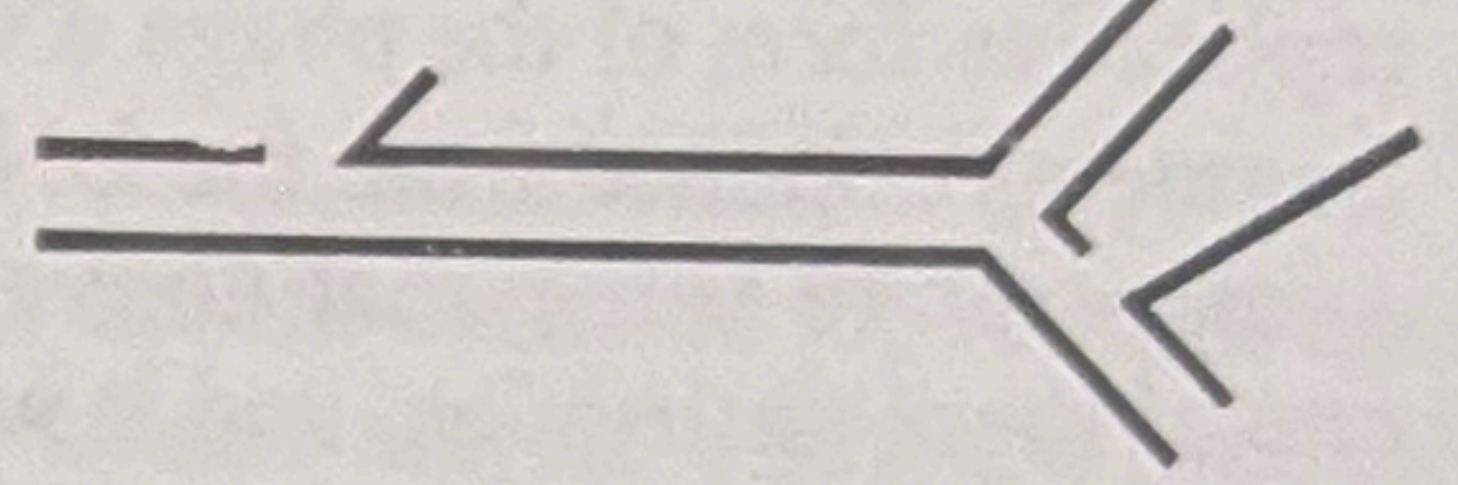


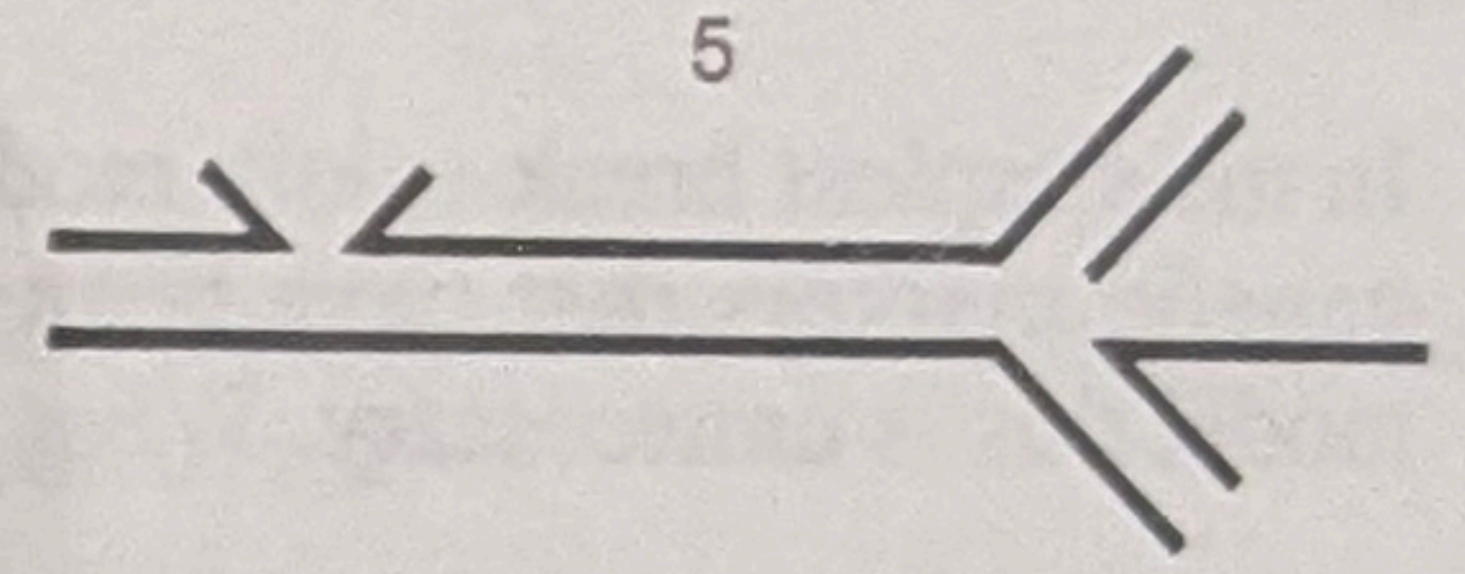
Original molecules

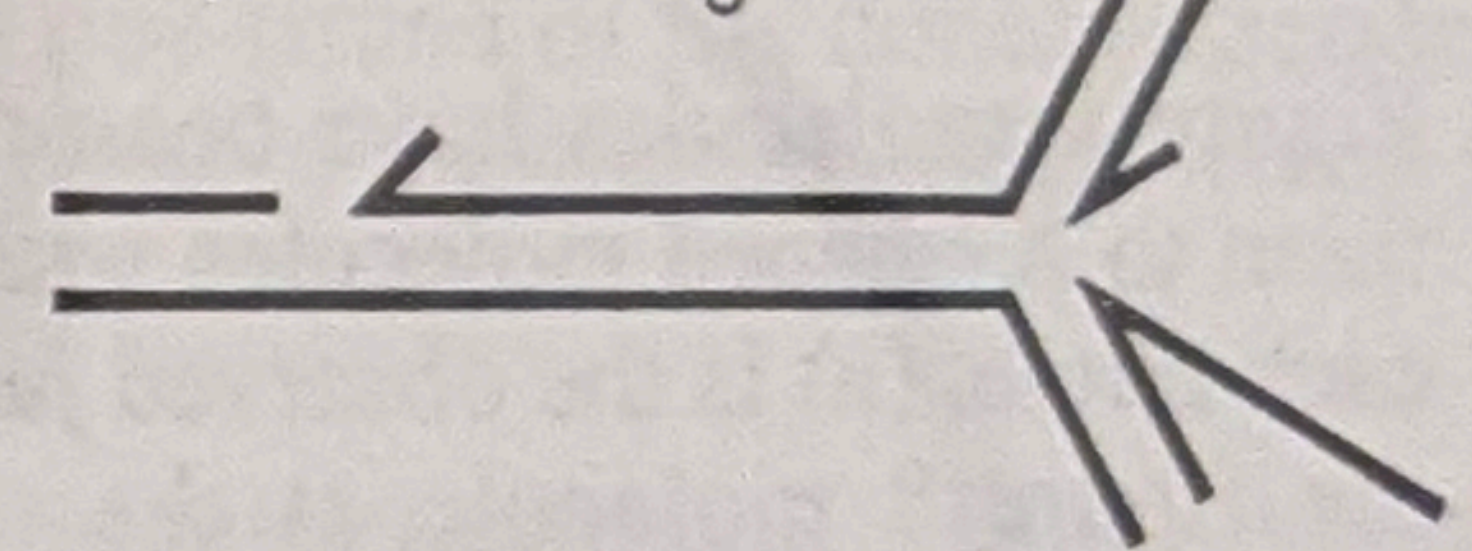
1


2


3


4


5


6


In each molecule a + strand is complementary to its own − strand and to the − strand of the other molecule. The molecules are oriented such that the *b* ends are homologous in base sequence. The upper molecule is shorter than the lower one and lacks the *a* end. The molecules are subjected to renaturing conditions during which they hybridize and branch-migrate. Which of the configurations, 1 to 6, can arise?

13 In the simplest breakage-reunion models the initial exchange must occur between the markers that have been recombined. In the asymmetric strand-transfer model this is not necessary. What feature of the model eliminates this requirement?

14 In the asymmetric strand-transfer model gene conversion can arise in two ways. What are they?

15 In *Pneumococcus* transformation is typically carried out by plating recipient cells, which have taken up DNA, directly on a selective growth medium that allows growth only of recombinants. Genetic markers in *Pneumococcus* fall into two classes, low-efficiency and high-efficiency markers. The latter, when present in the donor molecule, appear in the recombinants much more often than the low-efficiency markers do. There is no difference in the degradation of DNA molecules containing the two types of markers, and both are taken up equally well by competent cells. Suggest an explanation for the two classes.

16 Two double-stranded circular DNA molecules recombine according to the Holliday model. What is the form of the unit containing the Holliday junction?

17 Radding did an experiment in which both superhelical and nicked circles of the replicative form of the phage ϕX174 were separately mixed with single-stranded fragments obtained from the DNA of the phage. No RecA protein was added. At a temperature somewhat above room temperature strand invasion occurred in one of these molecules, and it was not possible to find a temperature at which it would occur in both molecules. Explain these results.

18 In the simplest break-rejoin models the initial exchange must occur between the genetic markers that have been recombined. In the asymmetric strand-transfer model this is unnecessary. What feature(s) of the model eliminate(s) this need?

19 What physical features define a Holliday intermediate?

20 Figure-8 molecules have been detected by electron microscopy by converting them to X-shaped molecules with a restriction enzyme that makes a single cut in each circle. (a) Is the observed junction probably at the same position as it was in the figure-8 molecule at the time of isolation from infected bacteria? (b) Figure-8 molecules are quite stable but if X molecules are stored for several hours before preparing them for electron microscopy, no X molecules are found; only linear molecules are seen. Explain this phenomenon.

ANSWERS TO PROBLEMS

1 Base-sequence homology and a RecA-type protein are required for transduction, transformation and bacterial conjugation. Meiotic recombination requires homology and presumably a RecA-type protein also.Transposition and site-specific exchange require neither. One way that transposition is usually detected is by observing the genetic change in a mutant having a nonfunctional RecA protein.

2 Prophage integration and excision, formation of an F′, formation of an Hfr, and meiotic recombination are reciprocal events, because neither genetic information nor DNA is lost. Meiotic recombination may appear nonreciprocal when conversion occurs, but the physical exchange remains reciprocal. Transformation is nonreciprocal, because one DNA strand is discarded and because only a portion of the strand that enters the bacterium is ever integrated. Transduction is nonreciprocal, because material from both ends of the double-stranded DNA fragment are discarded.

3 It increases recombination frequency by providing breaks, gaps, and free ends, which facilitates pairing. Also, synthesis of the RecA protein is stimulated.

4 White (nontransformant, or a cell in which the lac^+ DNA was integrated and in which there was a conversion from + to –), red (one lac^+ strand integrated and a conversion from – to +), and sectored (one lac^+ strand integrated and no conversion).

5 They would not be like transformation since a double-stranded DNA must enter the cell and in transformation only a single strand enters. They are probably similar to one another since double-stranded DNA is involved, two exchanges are needed, and RecA protein is required.

6 The Hfr cell is RecA$^+$, so the *recA* gene is also transferred at some time to the recipient. Once this happens, recombination can occur.

7 Postmeiotic segregation, for in these asci pairs of spores carry different alleles and each member of the pair is derived from one strand of the double-stranded meiotic product. Thus, each strand carries a copy of a different allele, which is one definition of a heteroduplex region.

8 I are nonrecombinant. II are recombinant without conversion. III are recombinant with an *E*-to-*e* conversion.

9 The existence of aberrant 4:4 asci shows that heteroduplex regions are produced in the exchange process. The remaining types all arise when mismatch repair of a heteroduplex regions occurs; since they are absent, the mismatch repair system must be absent. This situation can be observed with mutants that eliminate activity of the mismatch repair system.

10 Branch migration causes the base-paired region to drift back and forth. However, migration cannot occur across the nonhomologous region, so the single strand can never be completely assimilated. As the branch migrates back and forth, occasionally migration will extend to the homologous end of the single strand, at which point the single strand will dissociate completely from the circle. The single strand cannot penetrate the circle again at a low temperature, so ultimately all circles lose the partially assimilated strand.

11 For distant markers, in the resolution step there is a probability of 50% of cuts that would produce recombination and 50% of cuts that yield the nonrecombinant configuration of markers.

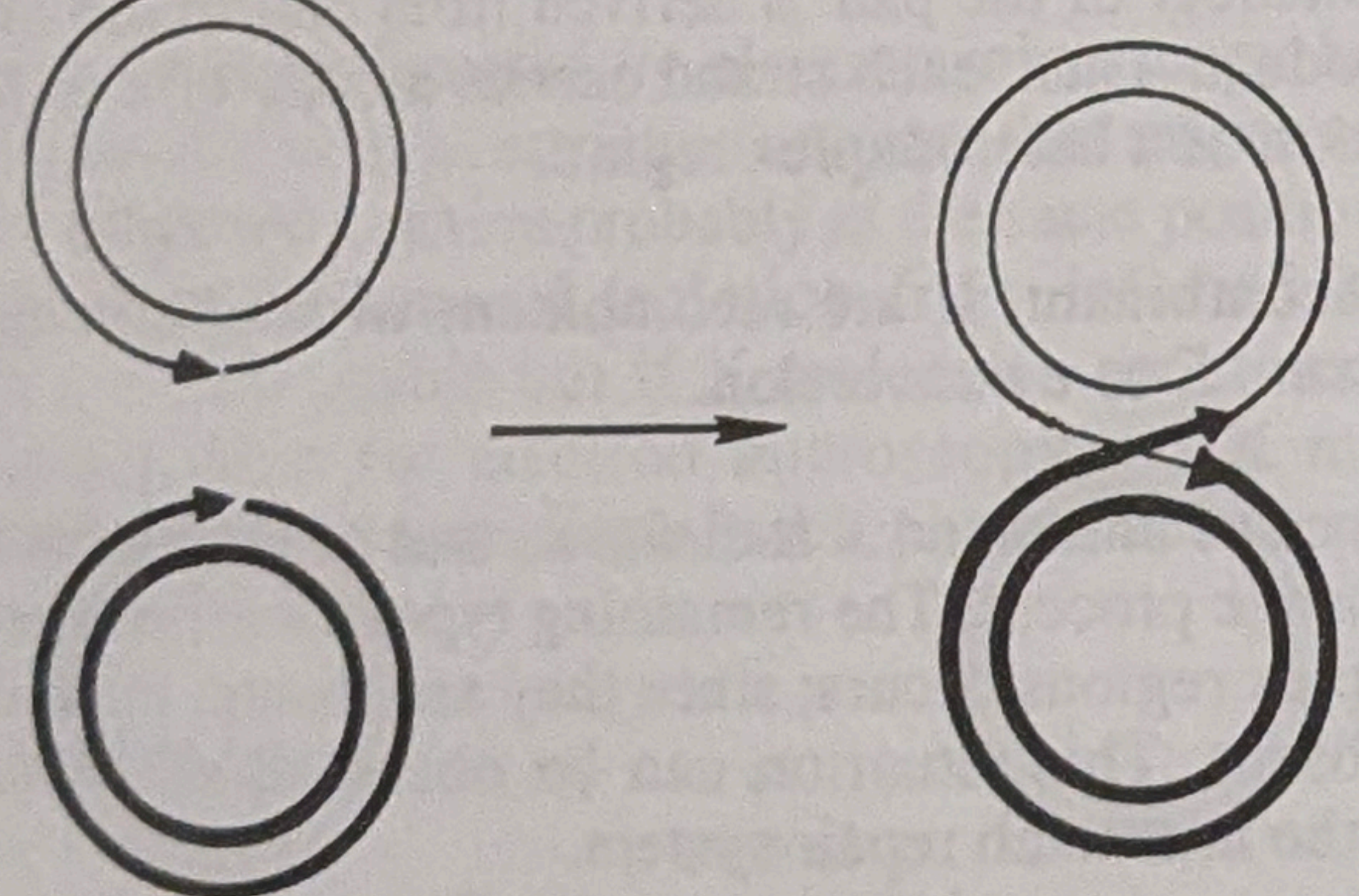
12 3 and 5. Molecule 1 is impossible because the two strands shown as forming a short double-stranded segment at the upper left have the same polarity and could not base pair. Molecule 2 is not possible because the short double-stranded segment in the upper part of the right fork consists of two segments having the same polarity. Molecule 4 has the same deficiency as molecule 2. In molecule 6 the short and long single-stranded appendages in the right fork would form a short double-stranded appendage, because they are adjacent and in register.

13 Branch migration, for the initial point of joining is not usually the point at which resolution of the Holliday junction occurs.

14 Mismatch repair and by copying one parental strand during assimilation.

15 Immediately after invasion, the recipient has a mismatched base pair. For low-efficiency markers mismatches are usually corrected to the recipient genotype. For high-efficiency markers either mismatch repair fails to occur or correction is primarily to the donor genotype.

16 Two circles joined by crossed strands, that is, a figure 8. The mechanism of formation of the figure-8 molecule is shown below.



17 Strand invasion occurred with the superhelix, for a naturally occurring (negative) superhelix has long-lived, somewhat long, single-stranded regions, owing to transient unwinding. Invasion fails with nicked circles because, if the temperature were high enough to maintain large single-stranded regions, it would be too high for stable binding of a single-stranded fragment.

18 Branch migration.

19 Two double-stranded DNA molecules connected by two nearby single strands.

20 (a) No, the junction will drift by branch migration. (b) The junction, moving back and forth by branch migration, ultimately reaches an end of the X molecule and the two double-stranded units separate.

MORE CHALLENGING PROBLEMS

1 Suppose that phage mutants are isolated that show decreased ability to recombine. All mutants fall into two complementation groups, *recX* and *recY*. The known genes of the phage have the map order *d e f g h i j k l m n*. Crosses are performed with Rec⁻ mutants. The data are shown below:

Genotypes		*Recombination frequency, %*	
Parental phages in crosses	*Recombinant*	*recX*⁻	*recY*⁻
d^-, e^-	++	0.0001	1
f^-, j^-	++	0.0001	4
j^-, m^-	++	1	2
j^-, n^-	++	1	3
k^-, m^-	++	1	1
k^-, l^-	++	1	0.01
j^-, k^-	++	0.0001	1
l^-, m^-	++	0.0001	1

(a) What properties of the RecX and RecY systems are indicated by these results? (b) What is the probable relation between the physical spacing of adjacent genes?

2 A λ phage can be crossed with a *gal*-transducing particle even though there is a region of nonhomology. For instance, there is no difficulty in detecting P^+Q^+

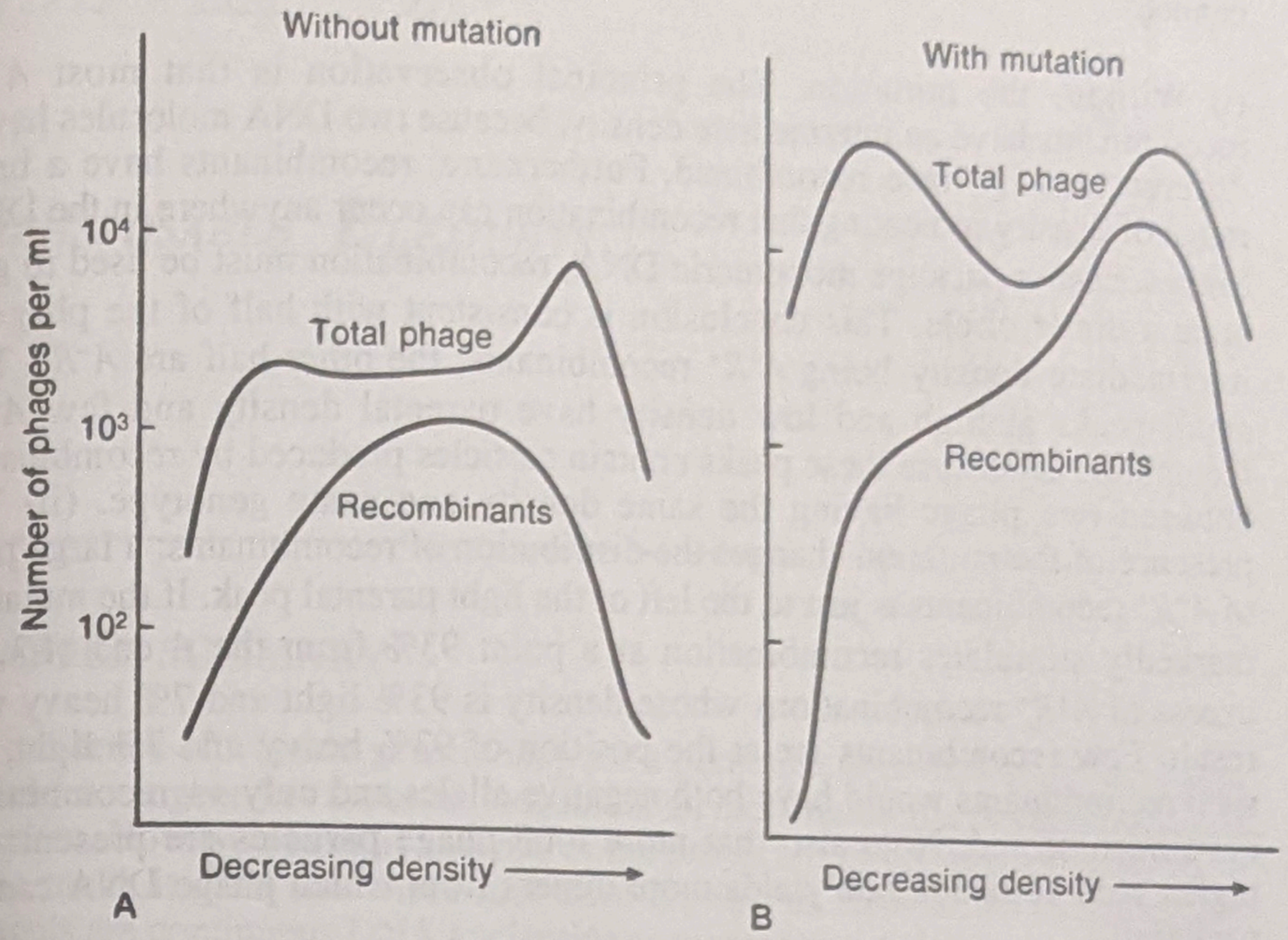
recombinants in a cross between a P^- *gal* particle and a Q^- normal phage. However, if two markers are very near but definitely not within the region of homology, the recombination frequency between them is depressed compared to that with normal phages. How would either Holliday or Meselson and Radding explain this difference?

3 A phage has a gene *a* and a gene *x* at opposite ends of its linear DNA molecule. A bacterial culture is infected with equal numbers of a^- and x^- phage particles using conditions in which DNA replication cannot occur. The average burst size of the phage is one, and 25% of the progeny have the genotype a^+x^+. What fraction of the input phage DNA molecules have engaged in genetic recombination?

4 In studying figure-8 molecules of phage ϕX174 by electron microscopy it is important to be sure that an observed figure-8 did not simply arise by two circular molecules overlapping one another. (a) What criteria must be applied to an observed structure before even considering that it might be a figure-8? (b) What could you do when preparing the DNA sample for electron microscopy to avoid the overlap problem?

5 *E. coli* strain Q is RecA$^-$ (that is, it is deficient in bacterial recombination) and contains a portion of a λ prophage—only the *J* gene and the right prophage-attachment site. The *J* gene is mutant and codes for the *h* character. Wild-type λ cannot adsorb to the λ-resistant strain K/λ, but a λ mutant carrying the *h* marker can. If a cI^- mutant is plated on a bacterial lawn consisting of a 1:3 mixture of Q and K/λ, plaques result which are easily seen, although they are somewhat turbid. Many phage recovered from the plaques can form plaques on K/λ. When a mutagenized stock of λcI^- is plated on this mixture, some plaques are found which are so turbid that they are almost not visible. These plaques do not contain phage that can form plaques on K/λ. What is the probable genotype of the phage producing these very turbid plaques?

6 In the following experiment all λ phage with genotype $red^-gam^-A^-$ were density labeled with ^{13}C and ^{15}N and were mixed with λ$red^-gam^-R^-$ containing ^{12}C and ^{14}N. The mixture was used to infect *E. coli* using conditions in which DNA replication was inhibited and in which each bacterium was infected by 10 phage particles. The phage that were produced were centrifuged to equilibrium in Cs formate and the distribution of genotypes throughout the density gradient was determined. The distribution of total phage and A^+R^+ recombinants shown in the figure at the top of the following page was observed. In a second experiment both phage contained a mutation known to map approximately 93% from the left end of the phage and the distribution shown in the figure was observed. Explain both results.



ANSWERS TO MORE CHALLENGING PROBLEMS

1 (a) *RecX* is a gene for general recombination; *recY* catalyzes site-specific exchange between genes *k* and *l*. (b) All genes are roughly equally spaced except *k* and *l*, which are very near.

2 Both would say the same thing. Branch migration cannot occur across the nonhomologous region, which limits the size of the region in which the initial association between the molecules can occur. Reducing the size of this region will reduce the probability of interaction and hence the recombination frequency.

3 If 25% are a^+x^+, then presumably 25% are a^-x^-. Since pairing is random, for every pairing event between a phage and another of unlike genotype, a pairing event must occur between two particles of like genotype. Since 50% of the phage are recombinant, 50% must also be products of parents of like genotype. Therefore, all of the DNA participated in recombinational events.

4 (a) It must be possible to draw the observed configuration in a way that two circles of equal size are joined at one point. (b) Adjust the DNA concentration to a very low value so that overlapping is very infrequent.

5 Red^-; red^+ phage can acquire the *h* marker by recombination, but these mutants

cannot.

6 (i) Without the mutation. The principal observation is that most A^+R^+ recombinants have an intermediate density, because two DNA molecules having different density have recombined. Furthermore, recombinants have a broad range of density indicating that recombination can occur anywhere in the DNA. Since λ cannot package monomeric DNA, recombination must be used to generate a dimer circle. This conclusion is consistent with half of the phage of intermediate density being A^+R^+ recombinants; the other half are A^-R^-. The small peaks at high and low density have parental density and few A^+R^+ recombinants because these peaks contain particles produced by recombination between two phage having the same density and same genotype. (ii) The presence of the mutation changes the distribution of recombinants: a large peak of A^+R^+ recombinants is just to the left of the light parental peak. If the mutation markedly stimulates recombination at a point 93% from the *A* end of λ, an excess of A^+R^+ recombinations whose density is 93% light and 7% heavy will result. Few recombinants are at the position of 93% heavy and 7% light, for such recombinants would have both negative alleles and only ++ recombinants have been scored. Note also that more total phage particles are present: the higher recombination rate yields more dimers from which phage DNA can be packaged.

CHAPTER 21

TRANSPOSABLE ELEMENTS

CHAPTER SUMMARY

The existence of a stable genetic map for a particular organism suggests that at all times each gene has the same location in all individuals of a population of that organism. However, careful analysis of the properties of certain unstable loci in maize made it clear that these loci were capable of movement. Transposable elements are continuous DNA nucleotide sequences that can relocate themselves from one part of the genome to another. In prokaryotes the transposition process is replicative in that relocation normally consists of placement of a copy of a transposable element at a new site, with another copy remaining at the original site. In eukaryotes transposition may also be replicative but can also apparently occur by excision of the element and reinsertion at a new site. The molecular mechanism of transposition is unknown. The essential features of the process are (1) the duplication of a short sequence (the target sequence, <10 base pairs long) surrounding the relocated element and (2) independence of a host *recA* or *recA*-like gene. The length of the target sequence, but not the sequence itself, is a characteristic of a particular transposable element. An enzyme encoded in the element, a transposase, probably mediates transposition in most cases. Transposable elements consist of several hundred to several thousand nucleotide pairs and have inverted-repeat termini 15–75 base pairs long. Composite transposable elements consist of a central segment of several thousand base pairs flanked by identical copies of smaller simple transposable elements at each end. Since the simple elements possess termini in inverted repeat, the composite elements also have inverted-repeat termini. In prokaryotes the number of copies of a particular transposable element is fairly small, usually <10; in eukaryotes, up to several hundred thousand copies of some transposable elements have been observed.

A variety of genetic events result from the presence of transposable elements. In bacteria some of these elements contain genes for antibiotic resistance. Others contain termination sites, which cause polar mutations, if the element is inserted between a promoter and a normal transcription signal.

Transposition can also occur within a gene; this completely eliminates gene function. Other events mediated by transposable elements are deletion, inversion, plasmid fusion, and replicon fusion. Transposition from a chromosome to a plasmid provides homologous regions between which RecA-mediated genetic recombination can occur; this can also lead to replicon fusion, as in the formation of *E. coli* Hfr strains from cells containing F. Another RecA-mediated event is a deletion of a segment between a pair of identical transposable elements, as described in Chapter 11.

Genetically, transposition is usually detected in bacteria by the movement of a gene for antibiotic resistance. In eukaryotes transposition is recognized by changes in gene expression. A nongenetic way to recognize a transposable element is to observe in a nucleotide sequence a short duplicated sequence (the target sequence) flanking a larger sequence.

Many models have been proposed to explain transposition, but none have been proved. It is unlikely that a single model will apply to all transposable elements. For example, some elements form a cointegrate, which contains two copies of the element, as an intermediate, but this is not true of all transposable elements. An interesting feature of transposition is that the number of copies of a particular transposable element is limited to a value characteristic of the element. No information is available about the mechanism of such limitation.

Eukaryotic transposable elements account for most, if not all, of the middle repetitive sequences of the higher eukaryotes. There are probably many classes of eukaryotic elements. Some exist both in integrated form and as free circular DNA molecules. Others, the retroposons, transpose via an RNA intermediate and utilize cellular or element-encoded reverse transcriptase to form the final DNA copy. The Alu sequences are thought to be retroposons that have evolved from tRNA molecules. Most of the retroviruses have many features in common with retroposons. The foldback elements are an interesting class, consisting of a sequence that is repeated numerous times within a particular element.

Although transposable elements are widespread and affect the expression of some genes, they are probably not essential. Some *Drosophila* species contain no detectable transposable elements. Most bacteria lack one or more of the known transposable elements, and deletion or duplication of any of these elements does not generally alter the well-being or growth characteristics of the cells. It is possible that transposable elements fall into the class of sequences occasionally called selfish DNA, sequences that exist because they have gained the ability to maintain themselves.

BOLD TERMS

Alu sequences, asymmetric models, cointegrate, FB elements, foldback elements, hot spots, insertion sequences, internal resolution site, IS elements, polar mutations, replicon fusion, resolvase, retroviruses, target sequence,

transposable elements, transposase, transposons, transposition

ADDITIONAL INFORMATION

In the text we showed how a transposon can be detected in a bacterium as the acquisition of an antibiotic-resistance gene by a plasmid from another plasmid. Another detection scheme uses a temperate phage. This method is of interest because it can be used to generate bacterial mutants. Consider a RecA$^-$ cell containing an *amp-r* R plasmid in which the *amp-r* gene is carried on a transposon. The cell is infected with a temperate phage, such as λ. About 1 progeny phage in 10^7 will acquire the transposon carrying the *amp-r* gene. The resulting phage population is used to infect an Amp-s (ampicillin-sensitive) culture of bacteria at a ratio of about 1 phage per 10 bacteria in a growth medium containing ampicillin. Uninfected bacteria are killed, and cells infected with a phage lacking the *amp-r* gene will not yield progeny phage because of premature lysis induced by the drug. However, if the phage carries the *amp-r* gene, the infected bacterium will survive and all progeny will carry the *amp-r* transposon. Proof that the *amp-r* gene is still part of a transposon comes from infecting an Amp-s RecA$^-$(λ) lysogen with the λ*amp-r* phage. Since the lysogen and the phage are homoimmune, the phage are unable to develop and the bacteria survive. The infected cells are then plated on agar containing ampicillin. The *amp-r* phage enables the cells to grow, but, since the phage cannot replicate (because of the presence of the phage repressor), in each division cycle one daughter cell does not inherit a λ*amp-r* DNA molecule, and a bacterial colony cannot form. However, in a few bacteria, transposition does occur, and the bacterial chromosome acquires the *amp-r* transposon, and all daughter cells do inherit an *amp-r* gene. Thus, formation of a colony indicates that transposition has occurred.

This procedure can be used to generate mutants of particular genes. For example, in the infection of the homoimmune cell the transposon might insert in the *lac* gene. Thus, if the Amp-s lysogen were *lac*$^+$ and a *lac*$^-$ mutant were desired, one would look among hundreds of Amp-r colonies for one that is Amp-r Lac$^-$. This could easily be done by plating the infected cells on a color-indicator agar containing lactose and ampicillin, on which Lac$^+$ and Lac$^-$ colonies can be distinguished by their different colors. This technique is a useful one for preparing defective mutants since the fraction of drug-resistant colonies that are mutated in a particular gene is generally higher than the fraction of the mutant cells induced by most mutagens. The mutants obtained are always absolute defectives, rather than conditional mutants, because the coding sequence is interrupted by the transposon.

Transposition of bacterial transposons is dependent on the presence of a transposase, presumably encoded in each transposon. Recently it has been observed that synthesis of the transposase of Tn10 is regulated by an antisense RNA (see Additional Information, Chapter 15). Antisense RNA is transcribed from the transposase gene itself, and the presence of this RNA species blocks translation of the transposase mRNA. Reduction in synthesis of the transposase prevents transposition. How synthesis of sense and antisense RNA is balanced is not yet known.

DRILL QUESTIONS

1 What two components are characteristic of transposable elements?

2 If you were determining the base sequence of a long DNA sequence, what features of the sequence would make you confident that a transposable element was present?

3 A particular transposable element is observed at five different locations. Will the flanking target sequences be the same at each site?

4 A transposable element is observed at several different sites in different bacterial strains. Each element is flanked by a pair of target sequences. What characteristic is common to these target sequences?

5 What is a typical length of a target sequence?

6 What is a composite transposon?

7 Does a composite transposon necessarily have terminal inverted repeats?

8 Are all drug-resistance genes located on transposons?

9 Why does the presence of an IS element in a particular genetic region often cause polarity?

10 What is the evidence that transposition in bacteria includes a replicative step?

11 Is it likely that DNA replication is involved in transposition in eukaryotes?

12 Repetitive sequences in eukaryotes are very often transposable elements. Do you think all eukaryotic repetitive sequences that exist in large number are transposable elements?

13 When transposition occurs in prokaryotes, two base sequences are duplicated. What are they?

14 When transposition occurs, is there any site selectivity with respect to the new site?

15 What term is used to describe a plasmid that results from transposon-mediated fusion of two plasmids, one of which contained a transposon? How many copies of the transposon are present in the fused molecule?

16 In the Tn3 system what enzyme is responsible for formation of a cointegrate?

17 In the Tn3 system what enzyme breaks down a cointegrate into two independent replicons? How many copies of the transposable element exist after action of this enzyme?

18 Name five genetic phenomena mediated by transposable elements.

19 Are transposable elements ever seen as free DNA molecules? If so, where and in what form?

20 What characterizes a retroposon?

21 Some eukaryotic transposable elements are present in enormous quantities. Roughly what fraction of the genome of *Drosophila* is *copia*? It is not important to know the exact number, but it is worth knowing the order of magnitude.

22 The Alu sequences are defined by their sensitivity to the restriction enzyme AluI. What makes one think that they may be transposable elements? What makes one think they may be retroposons?

23 What two features distinguish the *Drosophila* foldback elements from other transposable elements?

ANSWERS TO DRILL QUESTIONS

1 Terminal inverted repeats, and a gene encoding a transposase.

2 A unique sequence flanked by a sequence in direct or inverted repeated (the termini of the transposable element), the entire three-component unit flanked by a short sequence in direct repeat (the target sequence).

3 No. Each target sequence will be different.

4 The number of nucleotide pairs.

5 5 to 10 nucleotide pairs.

6 A transposable element consisting of a unique sequence flanked by two copies of a simple transposon, which can be in an inverted- or direct- repeat array.

7 Yes. Although the terminal simple transposons can be in direct repeat, these

transposons have inverted-repeat termini, so there will always be inverted terminal sequences.

8 No.

9 Many of the IS elements contain transcription termination sites. Hence if they are downstream from a promoter, whether within a gene or between genes of a polycistronic region, they will cause premature termination and hence loss of expression of downstream genes.

10 When transposition occurs, a copy of the transposable element appears at a new position without loss of the original element.

11 It is difficult to be sure, but from what we know now, it is not necessary, since transposition appears to involve excision and relocation. However, replication, of a repair type, may be required for re-insertion of an excised transposable element. Certainly, since the target sequence is duplicated, replication of at least this small segment is needed.

12 No, for example, the satellite DNA in telomeres and centromeres does not consist of transposable elements.

13 The transposon itself and the target sequence.

14 Yes, since within a given DNA segment, for example, the *E. coli lac* operon, certain sites are occupied more often than others. It is likely that transposition can occur to any target sequence, yet certain sequences are more likely than others.

15 The fused plasmid is a cointegrate; it contains two copies of the transposon.

16 The transposase.

17 The enzyme is a resolvase. There are two copies of the transposable element, as is always the case with transposition in bacteria.

18 Replicon fusion, deletion, inversion, transposition, and mutation.

19 Yes, but only in eukaryotes and for certain transposable elements. They appear as circular DNA molecules.

20 It has an RNA intermediate in the transposition process.

21 Roughly 3% of the genome.

22 The Alu sequences are flanked by a short sequence in direct repeat, like a target sequence of a transposable element. Homology with particular RNA molecules likens them to retroposons.

23 Their terminal inverted repeats are very long, amounting to about 1/3 of the total number of nucleotides; for a more typical transposable element this fraction would be less than 5%. Second, the terminal segments consist of a short (10-bp) sequence repeated about 100 times.

PROBLEMS

1 In homologous recombination two DNA molecules break and rejoin. If the event is physically reciprocal, the amount of DNA present is the same as that present before recombination; if the event is nonreciprocal, the total amount of DNA present decreases. What can be said about the relative amounts of DNA before and after transposition?

2 Transposition is a type of nonhomologous exchange, suggesting that matching of sequences is not part of the process. However, there seems to be some sequence effect in that transposition occurs to some sites more than to others. The mechanism of this effect is unknown. Suggest what type(s) of things might account for this effect.

3 It is easy to understand how insertion of a transposon can cause a mutation, but what feature of some transposons is responsible for genetic polarity?

4 You have isolated a segment of DNA having about 500 base pairs. Several genetic experiments have suggested but have not proved that the segment contains a transposon. What physical or chemical experiment would you do to test the hypothesis?

5 An Amp-r plasmid whose replication is temperature-sensitive is introduced into an Amp-s cell by the $CaCl_2$-transformation method. After growth for several generations 10^7 cells are plated on agar containing ampicillin at 42 C. Fifty colonies form. (a) Give three mechanisms which could explain the presence of these colonies. (b) Which mechanisms would not occur in a $RecA^-$ cell?

6 You have isolated two different insertion mutations in phage λ. How would you determine whether the two insertions are the same transposon?

7 A particular IS element has inserted into different positions in two λ phages. In one case the insertion is at position 55; in the other, at position 75 (both measured on a scale of 100 from the left end of the phage. The IS element has a length equal to 1% of the size of the phage DNA. One of these phage variants shows polar effects on downstream genes; the other does not. (a) Explain the

difference in the polar effects. (b) How could your explanation best be tested by electron microscopy?

8 Sometimes in a bacterial strain carrying two different plasmids, one of which contains a transposon, a single plasmid arises that has the genetic information of both plasmids. If the cell is RecA$^-$, this fusion can occur in two possible ways: (1) illegitimate recombination and transposon-mediated formation of a cointegrate or (2) some other form of nonhomologous recombination. (a) What information would be necessary to distinguish these mechanisms unambiguously? (b) What physical experiment could you do to obtain this information? (c) What biological experiment might you do to obtain this information?

9 Two bacterial genes (*chk* and *lat*) are very near one another. When *chk*$^-$ mutants are isolated (they arise spontaneously at a frequency of about 10^{-6} mutants per cell per generation), about 1% of these mutants are also *lat*$^-$. This is a surprisingly high frequency for spontaneous formation of double mutants. (a) Could you guess what might be the cause of this phenomenon? (b) Would a measurement of this frequency in a RecA$^-$ cell be informative? Explain. (c) Would studying the effect of chemical mutagens be informative? Explain.

10 It is sometimes desirable to remove a transposon from a bacterium. This is possible because excision of transposons occurs spontaneously, though at a very low frequency and by an unknown mechanism.) If a transposon were inserted in the *lac* gene, one would only have to obtain a Lac$^+$ revertant colony to obtain cells that had lost the transposon. In some genes, revertants are not detected in such a straightforward way; however, the following technique is successful for transposons carrying some antibiotic-resistance markers. Consider a culture of cells in which a transposon carrying the *tet-r* gene is present in gene *x* rendering the cells X$^-$. Tetracycline is added to a growing culture and a few minutes later penicillin is added. Penicillin only kills growing cells. The Tet-s cells are inhibited by the tetracycline; Tet-r cells grow in the presence of tetracycline and hence are killed by the penicillin. After many hours of growth the cells are plated on agar lacking tetracycline. At a low but easily detected frequency colonies form whose cells lack a functional *tet-r* gene. Describe three mechanisms for generating these Tet-s cells and indicate which ones result in transposon loss.

11 Comment on the frequency of transcription of transposable elements. For example, do you think that transposable elements are transcribed constitutively? There is no correct answer, but be reasonable.

ANSWERS TO PROBLEMS

1 The total amount of DNA always increases when transposition occurs, because of the duplication of the target sequence. In prokaryotes the transposon is also duplicated.

2 An effect of sequence on the probability of the first breakage event seems most likely.

3 The presence of a transcription stop signal or an in-phase stop codon in the transposon.

4 Heat denaturation followed by quick cooling and electron microscopic viewing might show a stem-and-loop structure. This would be suggestive but by no means would it be proof. Only base sequencing can answer the question for, if a transposon is present in the sequence, a target sequence must be duplicated. If a short duplication is seen, this is again only suggestive. If a duplication is not contained in the sequence, one can say with certainty that no transposon (of the types known) is present.

5 (a) Integration of the plasmid into the chromosome by homologous recombination; reversion to temperature-insensitivity; transposition of the *amp-r* gene from the plasmid to the chromosome. (b) Integration (if homology-dependent).

6 You could start by forming heteroduplexes with a phage DNA molecule lacking the insertion. Each insertion will produce a single-stranded loop. If the loops are not the same size, the transposons must be different. If the loops are the same size, the two transposons are probably the same. A better test would be to move one of the transposons (A) to a nonhomologous DNA molecule. Then the DNA containing transposon A can be renatured with DNA containing transposon B. Only if the transposons are the same will heteroduplexes form. Radioactive RNA could also be obtained by *in vitro* transcription of the DNA containing A, and this RNA could be used in a hybridization experiment with the DNA containing B. Successful hybridization would indicate that the transposons are the same.

7 (a) The IS elements are inserted in two different orientations and there is a transcription stop sequence in only one orientation. (b) Heteroduplexes between the DNA of the two phages would indicate that the transposons are the same or different.

8 (a) The number of copies of the transposon in the new plasmid would distinguish the mechanisms. This would be two, if the fusion was mediated by a transposon. (b) Form a hybrid with a nonhomologous molecule containing the transposon. If there are two copies of the transposon, a three-component structure consisting of one plasmid strand and two tester strands would be found. Base sequencing would of course also provide the information, but that would be much more tedious. Restriction enzyme analysis (described in Chapter 5) would be easiest. (c) Isolate the new plasmid and use the $CaCl_2$ technique to establish the plasmid in the Rec$^+$ cell. If there are two copies of the transposon, one component of the

hybrid plasmid should occasionally be lost by a homologous Rec-mediated recombinational event.

9 (a) A high frequency of adjacent mutations suggests that a deletion including these genes has occurred. The deletion frequency is sufficiently high that one might reasonably suspect that a transposon is adjacent to the genes and that the phenomenon is an example of transposon-mediated deletion. (b) Since many mutations result from SOS repair, which is absent in RecA$^-$ cells, and since the transposition is independent of RecA function, study of the phenomenon in a RecA$^-$ cell would be useful. If the deletion frequency were the same in Rec$^+$ and Rec$^-$ cells, a transposon would be implicated. (c) Most mutagens would increase the frequency of *chk*$^-$ mutations with little or no effect on deletion frequency, if transposition were involved. There might be an effect if the site represented a mutational hot spot and the phenotype was a result of a pair of point mutations.

10 (1) Precise excision of the transposon; (2) a deletion that either just removes part or all of the *tet* gene or the transposon plus adjacent DNA; (3) a mutation in the *tet* gene.

11 Transposition occurs so infrequently that it seems unlikely that transcription of most elements occurs very often. Since elements that regulate transcription have not been identified in transposons, it is possible that transcription is constitutive but weak. On the other hand, since the number of copies of a transposable element seems to be limited, it is possible that there is some type of repression system. In eukaryotes some transposable elements, for example, *copia*, are extensively transcribed.

MORE CHALLENGING PROBLEMS

1 Design a simple procedure for isolating the DNA of a transposon.

2 The frequency of reversion of transposon-induced polar mutations is very low. The frequency is higher in a RecA$^+$ cell than a RecA$^-$ cell, but the reason for this has not been unambiguously determined. Can you suggest a reason?

3 Since transposons are duplicated at a reasonably constant rate, one might expect that in the course of the millions of generations experienced by each bacterial species, a particular bacterial cell would contain thousands of copies of any transposon that it may have acquired in the distant past. However, one or two copies is the norm. (a) Explain this low number. (b) Suppose in nature a mutation had occurred which increased the transposition frequency 1000-fold compared to the frequencies usually observed. Would you expect such a transposon to exist today? (c) If the mutation had occurred in a laboratory strain, would you expect the situation to be different from that in part (b)?

ANSWERS TO MORE CHALLENGING PROBLEMS

1 Isolate two linear nonhomologous DNA molecules, each of which contain the transposon. These might be a plasmid DNA that had been cleaved to break the circle or a phage DNA. Mix the two DNA molecules, denature, and renature. Two types of molecules would result—perfectly renatured double strands obtained from the same from the same DNA molecule and heteroduplexes having one double-stranded region, which is the transposon. The single-stranded termini are digested with exonucleases. The full-length double-stranded molecules and the transposons can be separated by centrifugation or, better, by gel electrophoresis.

2 It seems possible that excision of a transposon could occur by means of RecA-mediated homologous recombination between duplicated target sequences. You may have other ideas.

3 (a) Because the size of genes is usually greater than the size of spacers and leaders, a great many mutations would be associated with having many transposons. Since each gene in a bacterial species serves a purpose, a large number of gene-inactivation events would surely weaken the cell and ultimately cause cell death. This probably is the reason that most transposons have been isolated from plasmids, since in general a plasmid is not essential to its host cell. (b) No, because the rapid accumulation of nonfunctional gene products, owing to repeated gene insertion, would lead to cell death. (c) Only very slightly. Such a mutant might survive somewhat better in the laboratory because very rich culture media are used; these media provide so many nutrients that more mutations can be tolerated than in nature.

CHAPTER 22

RECOMBINANT DNA AND GENETIC ENGINEERING

CHAPTER SUMMARY

The recombinant DNA technique allows different DNA molecules to be joined and propagated indefinitely, thereby creating novel genetic units. Restriction enzymes play an important role in the technique, because they are able to cleave DNA molecules within particular base sequences. Fragments obtained from any two DNA molecules cleaved by a single restriction enzyme can be joined by annealing their complementary single-stranded ends. The carrier DNA molecule used to propagate a desired DNA fragment is called a vector; the most common vectors are plasmids, phages, and viruses. The $CaCl_2$ transformation procedure is an essential stage in propagation of recombinant molecules, as it enables foreign DNA molecules to be taken up by bacteria. A variation of the technique is applicable to eukaryotic cells. If the recombinant molecule has its own replication system or is able to use the system of the host, it can replicate. If the molecule is a plasmid, it can become permanently established in the host cell. If it is a phage, it can multiply and produce a stable population of phage carrying foreign DNA.

Several methods are used for joining double-stranded DNA molecules: cohesive-end joining, blunt-end ligation, homopolymer tail-joining, and linkers. Cohesive-end joining uses the complementary single-stranded ends produced by most restriction enzymes. Two DNA fragments (one of which is typically a cleaved vector) obtained by cleavage with the same restriction enzyme have the same complementary termini and can be annealed together. Treatment with *E. coli* DNA ligase makes the joint permanent. Blunt-end ligation utilizes the ability of phage T4 DNA ligase to join molecules without single-stranded ends—for example, c-DNA, or those produced by certain restriction enzymes that yield blunt ends. In homopolymer tail-joining a single strand of poly(dA) is added to the 3' termini of a fragment, and a single strand of poly(dT) is added to the 3' termini of the cleaved vector. Then, the single strands (the tails) are allowed to anneal. Gaps are filled with DNA polymerase I and ligation seals the joint. The linker method enables one to join a molecule with blunt ends to a molecule with single-stranded termini produced by a restriction enzyme. In this method a short double-stranded segment, which contains the restriction site for the enzyme that yielded the molecule with cohesive ends, is blunt-end ligated to both ends of the fragment with blunt ends. The molecule with the linkers is then exposed to the restriction enzyme, which cleaves the linker and generates single-stranded ends that can be joined to the other fragment.

If a gene or a DNA fragment has been joined to a vector and the vector is in a form that can replicate (for example, as a plasmid, or in a phage), the gene or fragment has been cloned. A cell whose properties have been changed by establishing a recombinant DNA molecule is said to have been genetically engineered. The recombinant DNA technique has been used to clone genes, primarily for the purpose of obtaining large quantities of either a gene (for use as a hybridization probe or for nucleotide sequencing) or the gene product.

Several procedures have been used to obtain the original gene that is cloned. In some cases, the gene can be isolated from a restriction digest; occasionally, detection of the gene is straightforward, but sometimes ingenious and elegant methods, designed for the particular gene, are required. With eukaryotic genes a useful procedure is to isolate the mRNA derived from processing of the primary transcript. Such an RNA molecule can be converted to double-stranded DNA by the enzyme reverse transcriptase. Such DNA copies of an RNA molecule are called complementary DNA or c-DNA. They can be joined to various vectors either by homopolymer tail-joining or by the linker method.

A variety of procedures are used to insure that a plasmid vector contains inserted DNA. One method is insertional inactivation, in which an antibiotic-resistance gene is cleaved and insertion into the gene is allowed. If insertion has occurred, the plasmid will no longer be able to confer antibiotic resistance on the host cell. Typically, the plasmid contains two antibiotic-resistance genes. The gene for resistance to antibiotic #1 is cleaved and that for #2 is not. Thus, after joining and infection of antibiotic-sensitive bacteria, cells are sought which are resistant to antibiotic #2 and sensitive to #1. The cosmid technique also insures that a λ phage vector contains an insertion. In this technique ColE1 plasmid DNA containing one λ *cos* site is cleaved by a restriction enzyme and annealed with fragments obtained by cleavage with the same enzyme. The joined molecules are then subjected to a DNA packaging protocol. Only molecules containing two *cos* sites, which are obtained by annealing of two plasmids, are substrates of the packaging system. However, unless a DNA fragment has been inserted, it will be too small to be packaged. Thus, all viable phage will contain foreign DNA. Certain λ phage vectors contain the *E. coli lac* operon, and foreign DNA is inserted into the *lac* genes. Such insertion eliminates the ability to make β-galactosidase. Color-indicator agar can be used to determine whether the *lacZ* in the phage is active or inactive. Electrophoretic methods by which plasmid size is measured can also be used to identify a plasmid that contains foreign DNA, since the plasmid will become larger when DNA is incorporated.

Various techniques are used to identify a bacterial colony that contains a particular gene. Often a simple plating test can demonstrate the presence of the gene—for example, a test for utilization or production of a nutrient. However, for genes that do not produce a simply detected phenotype, other methods are needed. The presence of any gene can be detected by the *in situ* hybridization assay; in this assay the bacterial DNA is released from a colony on an agar surface, denatured *in situ* by flooding the surface with alkali, and radioactive mRNA, complementary to the DNA of the gene, is allowed to hybridize with the denatured DNA. If hybridization occurs, the colony becomes radioactive, and the radioactivity can be detected by autoradiography. Two immunological tests, the radioactive antibody test and the immunoprecipitation test, can also be used to detect an excreted protein if antibody to the protein is available. A variation of these immunological tests, in which cells are lysed by inducing a λ prophage with a thermosensitive repressor, enables one to detect proteins that are not excreted.

Cloning of a single strand of a fragment of double-stranded DNA is possible with the phage M13 cloning system. M13 contains a circular single-stranded DNA molecule but the replication cycle includes a double-stranded circular intermediate called RF. Initially cloning is accomplished by joining a double-stranded DNA fragment of interest to cleaved RF. Infection of *E. coli* with the RF produces phage particles containing only the phage (+) strand and hence only one strand of the cloned double-stranded fragment. The cloning region of the vector contains numerous restriction sites. If the RF is cleaved with two different enzymes, the two ends of the linear molecule will contain different cohesive termini. Cleavage of the double-stranded fragment to be cloned with the same enzymes yields a molecule that can be inserted in only one orientation into the RF vector. By appropriate choice of the cutting enzymes the fragment can be cloned in both directions. Thus, after infection of bacteria with the hybrid vector the phage will contain a single strand of the molecule to be cloned that depends on the orientation of the inserted double-stranded fragment.

Gene libraries are large sets of bacterial cultures or phage stocks, each of which contains one region of the genome of an organism. By appropriate choice of the number of cultures or phage stocks and the size of the cloned fragments, a library can be constructed that contains all regions of the genome of an organism of interest. For small organisms such as yeast, several thousand cultures are required to make a complete library. The number increases to tens of thousands (or higher) for the higher eukaryotes.

Synthesis of prokaryotic proteins in bacteria is straightforward, but synthesis of eukaryotic proteins by bacteria requires special procedures, such as linkage to prokaryotic promoters and ribosome binding sites. Eukaryotic proteins requiring posttranslational modification usually cannot be synthesized in bacteria, but some can be made in yeast. A few small eukaryotic proteins—for example, the polypeptide hormones—can be obtained by cloning synthetic DNA molecules, whose nucleotide sequence has been designed in accordance with the genetic code to yield a desired amino acid sequence. Specific vectors have been engineered to allow synthesis of eukaryotic proteins. A valuable type of vector is

the shuttle vector, an example of which is a vector that consists of an *E. coli* plasmid in which has been inserted a segment of yeast DNA containing selectable yeast genes, a yeast promoter, and a yeast replication origin. Cloning and selection can be carried out in yeast. However, the vector cannot be maintained in yeast since it lacks a centromere. Thus, the hybrid vector, once isolated, is transferred by the $CaCl_2$ transformation technique to *E. coli*, where it can be maintained. When the yeast product is desired, the shuttle vector can be transferred back to yeast. Mammalian expression vectors, consisting of bacterial and animal viral sequences, can also be used. The vector is maintained in *E. coli* but is transferred to an animal cell when gene expression is desired. Retroviruses and other viral vectors have been used to transfer genes into animal cells.

Genetic engineering has many applications outside of the field of biological research. For example, industrially important bacteria, capable of synthesizing economically valuable chemicals and certain drugs, are being prepared. Viral coat proteins synthesized in genetically engineered bacteria will certainly be a source of vaccines for viral diseases. Harmless viruses have also been modified to include the coat proteins of harmful viruses, again as a vaccine.

Procedures are being developed for modifying the properties of whole organisms, particularly humans, by infection with engineered retroviruses; this is called gene therapy. Whereas gene therapy has not yet been realized in any practical way and many problems exist, it is hoped that ultimately it can be used to restore missing genes to humans with particular diseases, such as diabetes.

A plasmid has been obtained from the bacterium *Agrobacterium tumefaciens*, which infects many plants, and this plasmid is being used to perform genetic engineering of higher plants. The goal is to modify plants by insertion of genes from other plants or organisms. An example is the construction of a nonleguminous plant that could fix nitrogen, thereby obviating the need for added fertilizer. Recently several plants have been created that are resistant to the herbicide glyphosate; this allows spraying of the herbicide on fields of growing plants so that weeds will be killed without harming the crop.

BOLD TERMS

alkaline phosphatase, c-DNA, cloned, cloning vehicle, colony, colony banks, complementary DNA, cosmids, gene libraries, gene therapy, genetic engineering, homopolymer tail joining, hybridization assay, immunoprecipitation test, insertional inactivation, linker, mammalian expression vector, polycloning site, precipitin, R-looping, radioactive antibody test, reverse transcriptase shuttle vector, subcloning, terminal nucleotidyl transferase, vector

ADDITIONAL INFORMATION

Chapter 22 dealt with methods for generating recombinant DNA molecules. A technique is also available for producing recombinant RNA molecules. It has not been used much but will probably become important in the coming years.

The *E. coli* phage Qβ is an RNA-containing phage that replicates its RNA by a phage-encoded enzyme called Qβ replicase. This enzyme makes a complementary copy of the template, and this copy is also a template for the enzyme. A second round of replication yields two more molecules, producing a total of four molecules, each of which is a template. The process continues, with the number of molecules increasing exponentially. An important point about the replication process is that as a product RNA chain is formed, it does not base-pair with its template, even though the base sequences are complementary. Such independence is clearly essential for the production of progeny phage RNA molecules and is probably a result of the large amount of internal base-pairing that occurs while each molecule is being synthesized. Qβ replicase is highly specific in that initiation requires particular base sequences; this is of course an efficient requirement for phage reproduction, to avoid rampant replication of rRNA, tRNA, and bacterial mRNA molecules. The recombinant RNA technology is based on the ability to insert an RNA molecule of interest into a shortened (221-base) Qβ RNA molecule (called midvariant RNA) capable of being copied by the replicase.

Two methods are used to join molecules to midvariant RNA. In the first, direct RNA joining, a recombinant midvariant containing a sequence of 10 adenines is used. This RNA is hybridized with decathymidylic acid, which provides a short double-stranded DNA-RNA hybrid region. This region can be cleaved by RNase H, an enzyme that cuts specifically in DNA-RNA duplexes. The thymidylic acid is then removed, yielding two RNA fragments, which are separated. The RNA to be inserted is joined by T4 RNA ligase with the 3' fragment and then this hybrid unit is joined to the 5' fragment, forming a recombinant RNA molecule. This recombinant is a template for the replicase. In joining the various fragments, appropriate removal and addition of phosphate groups is utilized to control the joining of desired fragments, as in the alkaline phosphatase method.

The second method makes use of the great manipulative power of the recombinant DNA technology. A DNA copy of midvariant RNA is made by the use of reverse transcriptase and this copy is inserted into a DNA plasmid with a promoter upstream from the insertion site. The "midvariant DNA" segment contains restriction sites that can be used to insert DNA molecules whose RNA copies are desired. Transcription of the plasmid with RNA polymerase yields a recombinant RNA molecule, which can then be used as a template for amplification by the replicase. With this method the desired RNA molecule, which might be difficult to isolate, is never needed in constructing the template.

DRILL QUESTIONS

1 Define cloning, vector, and vehicle as they are used in the recombinant DNA technology.

2 What properties of a DNA molecule are essential for it to serve as a vector?

3 How is recombinant plasmid DNA introduced into bacterial cells ?

4 Restriction enzymes generate three types of termini. What are they?

5 With the normal convention for orienting molecules, will a molecule with a 3' extension have the cut in the upper strand to the left or to the right of the cut in the lower strand?

6 Are the termini of a restriction fragment produced by a particular enzyme always the same or can they be different?

7 Can a restriction enzyme cut at two sites that have different base sequences?

8 Can two different restriction enzymes act at the same site?

9 What features do all restriction sites have?

10 Will the sequences 5'GGCC and 3'GGCC be cut by the same restriction enzyme?

11 Where do restriction enzymes come from?

12 What property must two different restriction enzymes possess if they yield identical patterns of breakage?

13 What unique property is possessed by the enzyme terminal nucleotidyl transferase, with respect to its polymerizing ability?

14 What three methods can be used to join fragments with blunt ends?

15 How must a blunt-ended molecule be treated before homopolymer tail-joining can be carried out?

16 In homopolymer tail joining, after annealing of a fragment and a vector it is invariably necessary to treat the joined molecule with DNA polymerase I? Why?

17 Name two methods that will ensure that a fragment will be joined to a vector in a unique direction.

18 How can one ensure that a plasmid vector does not re-form after cleavage with a restriction enzyme?

19 If a vector is cleaved with a restricion enzyme and then allowed to anneal

without exposure to any other restriction fragments, will any molecules form other than the original vector?

20 If a cleaved plasmid vector and a set of restriction fragments are annealed, some circular molecules are found that are smaller than the plasmid vector. Account for these molecules.

21 A bacterial gene is cloned in a vector next to a promoter. Hybridization with a radioactive mRNA probe indicates that the gene is present in the cell; However, the gene product is not made. What is the most probable reason for the lack of synthesis of the gene product?

22 Give several reasons why a cloned prokaryotic gene might not be expressed in a prokaryote. Assume that the vector does not provide any particular means for expression of cloned genes.

23 What method may be used to avoid the problems in Drill Question 22?

24 What is the advantage of using a plasmid with two antibiotic-resistance genes as a cloning vehicle?

25 What two methods are used to insure that a particular fragment has been cloned?

26 What advantage results from the use of a cosmid?

27 What feature of the M13 cloning system allows the cloning of a single strand of DNA?

28 Give two reasons why a simply cloned eukaryotic gene will not usually yield functional mRNA in a bacterial host.

29 Assuming the problems in Drill Question 28 are solved, give three reasons why a desired protein may not be made by a eukaryotic gene cloned in a bacterium.

30 Of what use is *Agrobacterium tumefaciens* in genetic engineering?

31 What element is usually lacking in a yeast-*E. coli* shuttle vector that requires that the plasmid be maintained in *E. coli*?

32 When a yeast-*E. coli* shuttle vector is used, is the initial selection by transformation carried out in yeast or in *E. coli*, and why?

ANSWERS TO DRILL QUESTIONS

1 Cloning is the act of inserting a particular gene or set of genes into a DNA whose replication can be controlled in order to produce many identical and easily recoverable copies of the gene or the set. Vector and vehicle have the identical meaning, namely, a DNA molecule into which a gene to be cloned can be inserted.

2 The vehicle must contain at least one suitable restriction site into which foreign DNA can be inserted without destroying the ability of the plasmid to replicate. Ideally, the vehicle should also have some selectable markers and, even better, some marker that changes as a result of inserting DNA at the restriction site. If the latter is the case, cells carrying the recombinant plasmid can be quickly selected and identified.

3 By the $CaCl_2$ transformation technique.

4 Blunt ends, cohesive ends with 3' extensions, and cohesive ends with 5' extensions.

5 To the right.

6 Always the same.

7 No. There is a class of restriction enzyme that recognizes a particular site but makes cuts at random positions around the site; these enzymes are not used in genetic engineering.

8 Yes. Several restriction enzymes from different organisms are known that have the same site specificity.

9 They are palindromes.

10 No.

11 They can be isolated from most, if not all, bacterial species.

12 They must both recognize the same base sequence.

13 It can add nucleotides to an extended single-stranded 3' terminus of a DNA molecule without the need of a template.

14 Blunt-end ligation, addition of linkers, homopolymer tail-joining.

15 Terminal nucleotidyl transferase will not add the tail to the blunt end. First, the DNA must be treated with a 5'-P-specific nuclease to remove a few nucleotides from the 5' end, leaving a 3' overhang.

16 The lengths of the homopolymer tails are not the same. Thus, when the

fragments are joined, there are either gaps or extra material in the joints. The gaps must be filled in before DNA ligase can seal the joint, and pol I can fill gaps.

17 Prepare fragments by cleavage with two different restriction enzymes such that their termini are different; put different linkers on a blunt-ended molecule.

18 Treat the cleaved plasmid with alkaline phosphatase. After joining and treatment with DNA ligase, heat the molecules; treated vectors will melt open.

19 Yes, dimers and higher multimers of the vector.

20 These are circularized fragments.

21 It has been inserted in the incorrect orientation.

22 The gene may have been separated from its promoter or from its ribosome binding site.

23 A standard procedure is to design a vector so that it contains a strong promoter and a ribosome binding site (commonly that for the *E. coli lac* operon), with an adjacent downstream restriction site.

24 One gene can be used to detect the plasmid in a transformation experiment; and if there is a restriction site in the other gene, lack of resistance to that antibiotic can be used to show that insertion has occurred.

25 A particular fragment is isolated from a gel after electrophoretic separation, or a particular c-DNA is used.

26 The transducing phage produced by the joining techniques must contain inserted DNA.

27 M13 phage contains single-stranded DNA but has a double-stranded replication intermediate into which a fragment of interest can be cloned.

28 (1) The bacterium does not recognize a eukaryotic promoter. (2) If a primary transcript is actually made, it will typically have introns and these cannot be removed in a bacterium.

29 (1) The mRNA lacks a ribosome binding site. (2) The protein must be processed. (3) The bacterium destroys the protein, having recognized it as foreign protein.

30 *Agrobacterium tumefaciens* contains a plasmid that can be used to perform genetic engineering of plants. This should markedly improve the techniques of plant breeding.

31 A centromere, without which progeny molecules would not always find their way into daughter cells.

32 In yeast. These shuttle vectors are used for cloning eukaryotic genes, which will usually not be expressed in *E. coli*, but may be expressed in yeast. Selection must be carried out in a cell in which the gene is expressed, that is, yeast.

PROBLEMS

1 Restriction enzymes A and B are used to cut the DNA of phages T5 and T7, respectively. A particular T5 fragment is mixed with a particular T7 fragment and a circle forms that can be sealed by DNA ligase. What can you conclude about these enzymes?

2 T7 DNA is digested with the EcoRI enzyme. Can all fragments circularize?

3 When DNA is analyzed by treatment with a restriction endonuclease followed by gel electrophoresis, it is almost always necessary to heat the mixture to 65 C before layering it on the gel. If this is not done, extra bands are seen. Why is this the case?

4 Is it possible for two restriction enzymes (acting separately on two DNA samples) to yield sets of fragments that are different but have one fragment whose size (number of nucleotides) in one set is identical to the size of a fragment in the second set?

5 Phage X82 DNA is cleaved into six fragments by the enzyme BglI. A mutant is isolated whose plaques look quite different from the wild-type plaque. DNA isolated from the mutant is cleaved into only five fragments. Account for this change. In general, discuss why often a restriction enzyme may be found that yields n fragments with wild-type phage and $n-1$ fragments with a mutant.

6 By using a restriction enzyme that generates blunt ends, a DNA sequence has been cloned in plasmid pBR322. The sequence was isolated and terminal nucleotidyl transferase was used to put a tract of cytosines at each end of the molecule. The cleaved plasmid had poly(dG) tails, and homopolymer tail joining was used to link the fragment to the plasmid. Several months later a new and uncharacterized restriction enzyme is found that will cleave the cloned sequence from the plasmid. What property must this enzyme have?

7 The enzyme terminal transferase is a DNA polymerase in that it adds a nucleotide 5'-P to a 3'-OH group. What property does it have that the usual polymerases do not have? Explain how it is used to join together two DNA molecules.

8 What two methods can be used to join blunt-ended fragments?

9 A large circular DNA plasmid is digested with a restriction endonuclease. The fragments are allowed to reassemble at random (that is, the fragments are no longer in the original order) and circles having the same molecular weight as the original plasmid are selected. Host cells lacking the plasmid are then infected with these fragments. (a) Assuming that ability to replicate is the only requirement for successful infection, what alterations in fragment order might prevent successful infection? (b) Consider a larger circle in which the order of the fragments is the same, but in which some circles contain an additional fragment adjacent to an identical fragment. Will this circle be capable of successful infection? (c) Is it possible to delete a fragment, keeping the order otherwise the same, and still infect cells successfully? Explain.

10 The *A*$^+$ allele of a cell is easily selected by growth on a medium lacking substance A. However, repeated attempts to clone the *A* gene by digestion of cellular DNA and the vector by the enzyme EcoRI (the vector has an EcoRI site) are unsuccessful. If HaeIII is used (the vector also has an HaeIII site), clones are easily found. Explain.

11 In cohesive-end joining why is DNA ligase needed to complete a joining reaction? In other words, since annealing has already joined the fragments and since replication can occur across a single-strand break, why must the joint be made a covalent one?

12 Suppose you had a plasmid with an *amp-r* gene having an EcoRI site and a *lacZ* gene with a SalI site. If you were trying to clone a particular gene, would it be easier to use EcoRI or SalI to cleave the DNA? Think about how you would detect the recombinant plasmid.

13 A Kan-r Tet-r plasmid is treated with the BglI enzyme, which cleaves the *kan* (kanamycin) gene. The DNA is annealed with a BglI digest of *Neurospora* DNA and then used to transform *E. coli*. (a) What antibiotic would you put in the growth medium to ensure that a colony has the plasmid? (b) What antibiotic-resistance phenotypes will be found among the colonies? (c) Which phenotype will have the *Neurospora* DNA?

14 A *lac*$^+$*tet-r* plasmid is cleaved by a restriction enzyme in the *lac* gene. The restriction site for this enzyme contains four bases and the cuts generate a two-base single-stranded end. The particular site in the *lac* gene is in the first amino acid of the chain, changes of which rarely generate a mutant. The single-stranded ends are converted to blunt ends with DNA polymerase I and then the ends are joined by blunt-end ligation. A *lac*$^-$*tet-s* bacterium is transformed with the DNA and Tet-r bacteria are selected. What will be the Lac phenotype of the colonies?

15 Reverse transcriptase, like most enzymes that make DNA, requires a primer. Explain why, when c-DNA is to be made for the purpose of cloning eukaryotic genes, a convenient primer is a short sequence of poly(dT).

16 A Tet-r plasmid having a single EcoRI site is cleaved by the EcoRI enzyme. The sample is annealed, treated with DNA ligase, and then used in a $CaCl_2$ transformation experiment. Tet-r colonies are selected. Gel electrophoresis of the plasmid DNA from these colonies shows that there are two types of colonies, one type yielding a single band and the other type yielding a band moving much more slowly. What might the slow band represent?

17 Consider a restriction enzyme that makes *x* cleavage sites within the *lac* operon, generating fragments containing (1) the promoter and operator, (2) the *lacZ* gene, and (3) the *lacY* gene. This collection of fragments is annealed with a vector and a *lacZ*$^-$*lacY*$^-$ cell is transformed. Among the Lac$^+$ transformants some clones make more β-galactosidase than permease and others make more permease than β-galactosidase. Give two different explanations for these clones. (Think about the fact that annealing is a random process.)

18 A large circular plasmid is digested with a restriction endonuclease. The fragments are allowed to reassemble at random (that is, the fragments are no longer in the original order and circles having the same molecular weight as the original plasmid are selected. Host cells lacking the plasmid are then infected with these fragments. (a) Assuming that ability to replicate is the only requirement for successful infection, what alterations in fragment order might prevent successful infection? (b) Consider a larger circle in which the order of the fragments is the same, but in which some circles contain an additional fragment adjacent to an identical fragment. Will this circle be capable of successful infection? (c) Is it possible to delete a fragment, keeping the order otherwise the same, and still infect cells successfully? Explain.

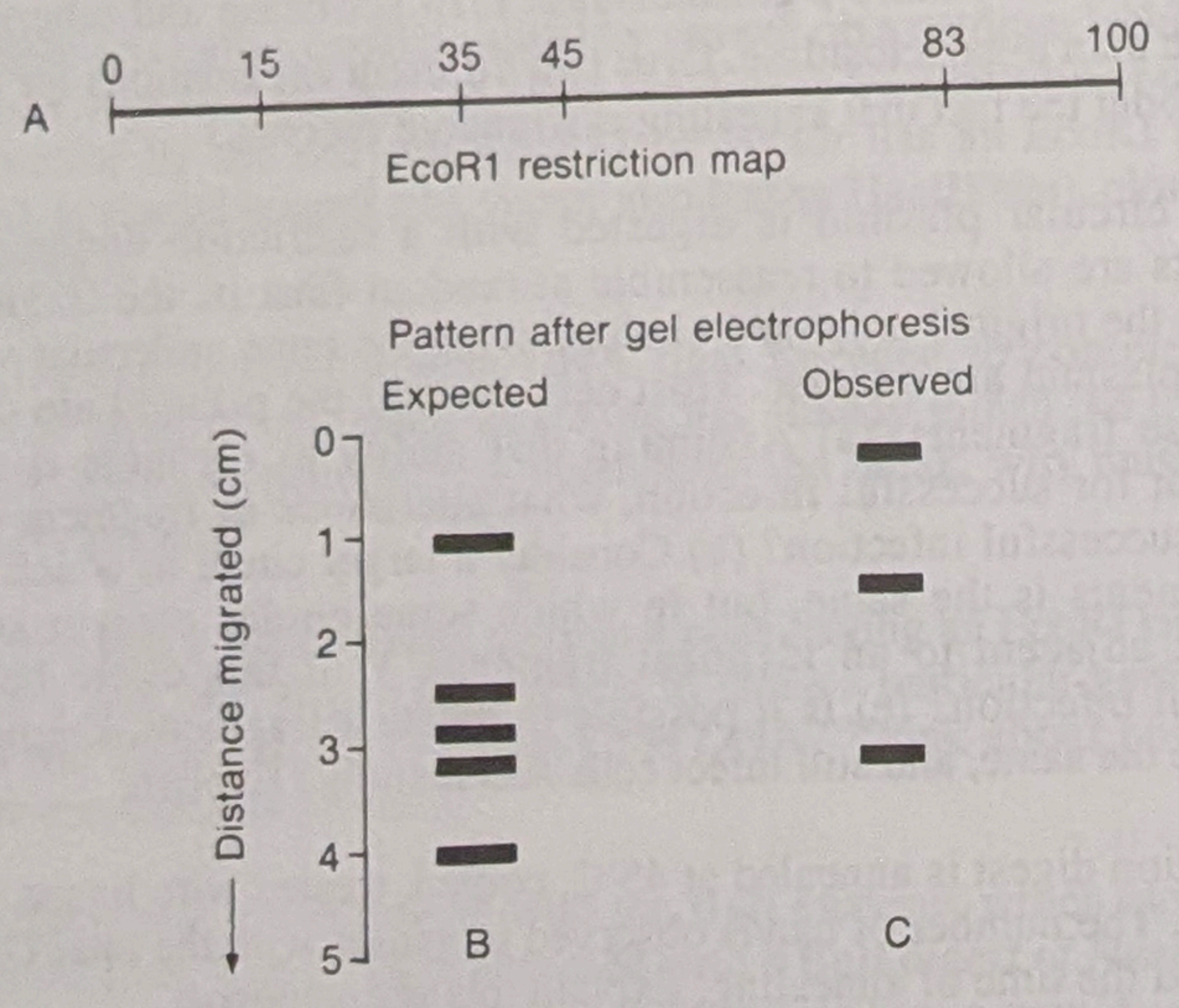
19 A restriction digest is annealed at 45 C, cooled, treated with ligase, and electrophoresed. The number of bands observed increases with the total DNA concentration and the time of annealing. Explain this phenomenon.

20 The DNA of temperate phage P4 is linear, double-stranded, 11.5-kb long, and has cohesive ends. Digestion with the BamHI restriction endonuclease yields fragments 6.4, 4.1 and 1.0 kb in length. Partial digestion with the BamHI enzyme yields fragments 10.5, 7.4, 6.4, 4.1, and 1.0 kb in length. Circular P4 DNA made with DNA ligase can be digested with the BamHI enzyme to yield fragments 6.4 and 5.1 kb in length. (a) What is the order of fragments in the DNA? (b) DNA is extracted from cells lysogenic for P4 and digested completely with the BamHI enzyme. A Southern blot is prepared with radioactive P4 phage DNA as a probe. Fragments whose lengths are 15.0, 12.5, and 6.4 kb are labeled with radioactivity. What conclusions can you draw from this experiment about prophage structure?

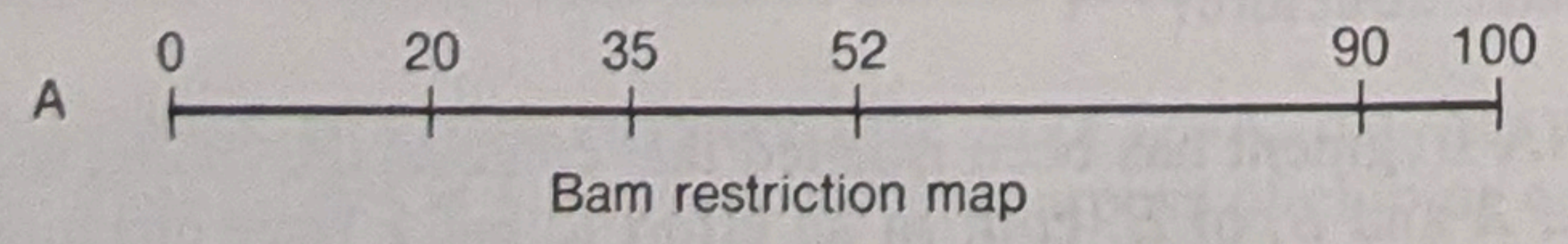
21 A DNA fragment has been isolated that contains the coding sequences of two genes, *A* and *B*, of *E. coli*. In an effort to make large quantities of the gene

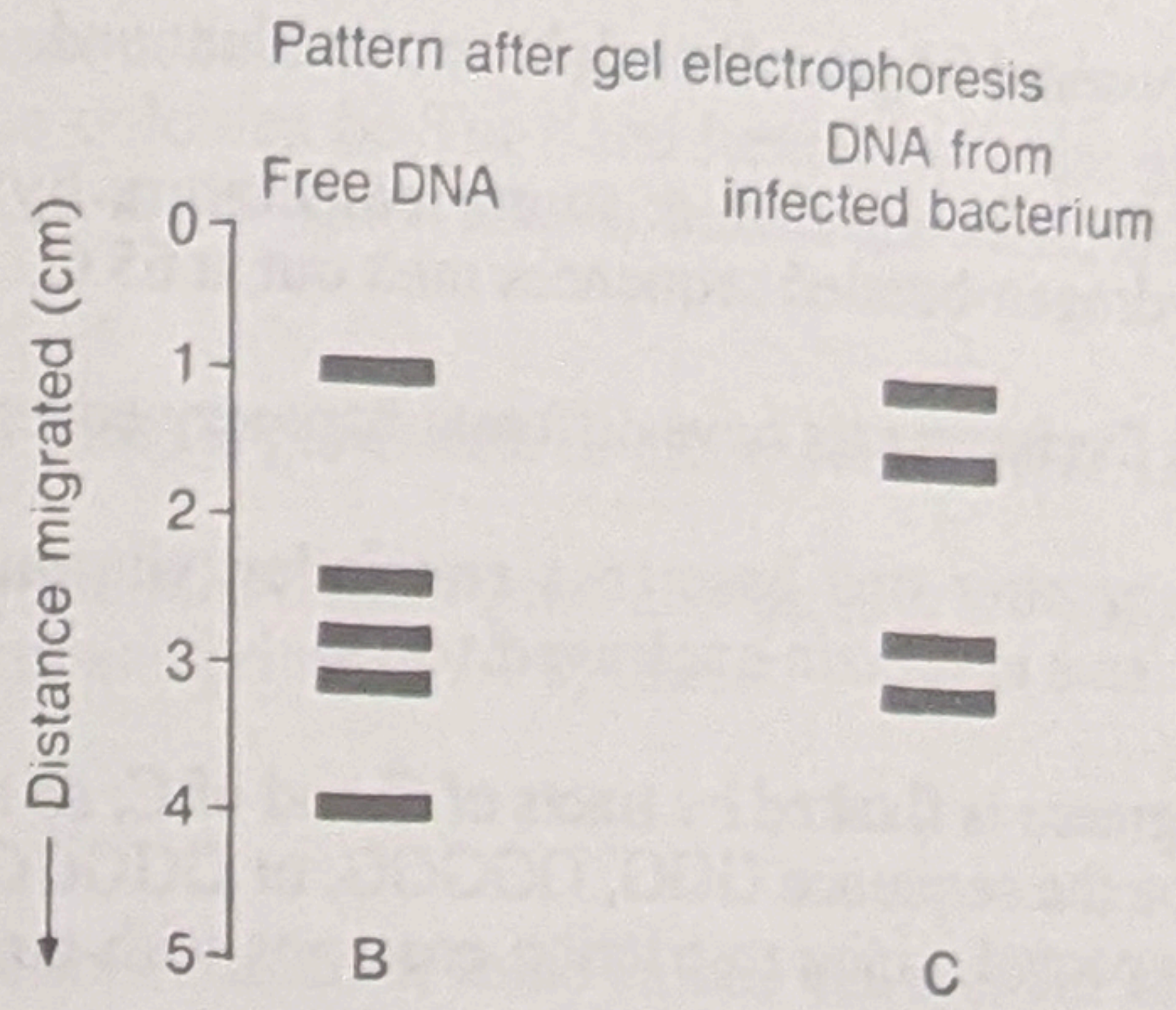
products, the fragment is joined to a standard plasmid and the DNA is used to transform an A^-B^- bacterium. Colonies containing the plasmid are isolated easily, but none are A^+B^+. However, A^+B^- and A^-B^+ colonies are found. Explain this observation.

22 A DNA molecule has the restriction map shown below when digested by the EcoRI restriction endonuclease (numbers refer to relative distance of each cut from the left end of the molecule). In a particular experiment (after heating the digested sample to 65 C), gel electrophoresis shows that instead of the expected five bands, the second pattern shown is observed. Explain.



23 A phage DNA molecule has short, complementary, single-stranded ends and circularizes when the phage infects a bacterium. The restriction map and the gel band pattern obtained after treatment of a free DNA molecule with the BamHI restriction endonuclease are shown at A and B in the figure below and at the top of the following page. You are studying the properties of DNA isolated from an infected bacterium. Under certain conditions the pattern of fragments obtained after enzymatic digestion and gel electrophoresis is that shown in panel C. What is the structure of the DNA under these conditions?




Pattern after gel electrophoresis
Free DNA
DNA from infected bacterium
Distance migrated (cm)
0
1
2
3
4
5
B
C

24 Sometimes a cloned bacterial gene is not expressed because it has been separated from its promoter. The gene can be expressed if it can be cloned in a plasmid containing a promoter to which the gene can be linked. The gene must be oriented correctly though, because the coding sequence must be linked to the strand containing the promoter. Unfortunately, when mixing the fragments with a cleaved plasmid, joining will occur in all possible orientations. Consider a plasmid containing the *E. coli lac* operon, with a BamHI site between the promoter and the genes. The *lac* sequence also contains restriction sites for other enzymes. Explain how any one of these enzymes might be used in the cloning procedure to ensure that the gene of interest is linked to the *lac* promoter.

25 High-copy-number plasmids are useful vectors for cloning genes whose gene product (the protein) is required in large quantities, for example, for purifying an enzyme that normally exists in very small amounts. A cell containing such a plasmid with a cloned gene may make enormous quantities of the desired protein. However, often a cell making a high concentration of an enzyme that normally exists in low concentrations cannot tolerate the high concentration, and the cell dies. Design a simple method for cloning such a gene in a high-copy-number plasmid.

26 How many clones are needed to establish a gene library for monkey DNA (3 x 10^9 base pairs per haploid genome) if fragments whose sizes average 2 x 10^4 base pairs are used and if one wishes 99% of the monkey genes to be in the library?

ANSWERS TO PROBLEMS

1 Both enzymes recognize the same base sequence. They might even be the same

enzyme.

2 No. The two terminal fragments each have one blunt end.

3 Fragments often join together at room temperature by their cohesive ends. These short hydrogen-bonded sequences melt out at 65 C.

4 Yes, as long as the fragments have different base sequences.

5 A mutation may alter one base in a restriction site and thereby cause two potential fragments to remain uncleaved.

6 The cloned fragment is flanked by tracts of G and of C, so the cleavage site of the enzyme must be the sequence GGG, GGGGG, or GGGGGG. Longer sequences of G are not expected, since restriction enzymes with target sequences greater than about six nucleotide pairs are rare.

7 Terminal transferase does not need a template. A poly(dA) strand can be added to one DNA molecule and a poly(dT) strand to another. The two molecules can be joined by hydrogen bonding between the poly(dA) and the poly(dT) tails.

8 Blunt-ended fragments can be attached to one another by homopolymer tail joining with poly(dA) and poly(dT) or homopolymer tail joining coupled with linkers having restriction sites.

9 (a) Successful infection would be prevented by an alteration that would split a gene whose product is needed for replication and a rearrangement that would break the sequence in the replication origin. (b) Yes, as long as it does not introduce the changes described in (a). (c) The answer is the same as in (b).

10 The *A* gene probably contains an EcoRI site, so the gene is destroyed. It does not contain an HaeIII site.

11 The cohesive ends are very short, so breathing will normally cause them to come apart at room temperature. In fact, the annealing reaction must sometimes be done below room temperature. The joint would not survive the higher temperatures (for example, 37 C) needed for transformation unless they were covalently joined. Without maintaining circularity, the DNA would be unable to complete its first round of replication within the host cell.

12 In principle, either could be used. If you used EcoRI, the plasmid could be detected by growth on medium using lactose as the sole carbon source. Colonies that grow could then be tested for antibiotic-sensitivity to find those containing DNA inserted in the *amp-r* gene. Note that this process would take two steps, and, in fact, two days of growth. If you used SalI, transformants with inserted DNA could be isolated in a single step by plating transformed bacteria on a lactose color-indicator medium containing ampicillin.

13 (a) The *tet-r* gene has not been cleaved, so addition of tetracycline to the medium will require that the colonies be Tet-r and hence carry the plasmid. (b) Tet-r Kan-r and Tet-r Kan-s. (c) Tet-r Kan-s, because insertion will occur in the cleaved *kan* gene.

14 A frameshift of two bases is generated, so all colonies will be Lac$^-$.

15 A processed eukaryotic mRNA will generally possess a 3' poly(A) tail. Poly(dT) will anneal to this tail and prime DNA synthesis.

16 A plasmid dimer.

17 Since annealing occurs at random, some clones will contain more than one copy of the *lacZ* gene and one copy of *lacY*, and others will contain more than one copy of *lacY* and one copy of *lacZ*. Furthermore, some will have the gene order, *lacP lacZ lacY* and others the gene order *lacP lacY lacZ*. Because of the polarity of synthesis of the gene products of a polycistronic mRNA, the former will make more *lacZ* and the latter will make more *lacY*.

18 (a) Successful infection would be prevented by an alteration that would split a gene whose product is needed for replication and a rearrangement that would break the sequence in the replication origin. (b) Yes, as long as it does not introduce the changes described in part (a). (c) The answer is the same as in (b).

19 At low concentrations circular monomers are the principal products. At higher concentrations dimeric, trimeric, etc., products form. At early times the number of components per annealed unit is smaller than at later times.

20 (a) The order is 1.0–6.4–4.1. The reasoning is the following. First, in the linear DNA the 6.4-kb fragment can be linked to either of the other fragments, when digestion with BamHI is partial; therefore, the 6.4-kb fragment must be in the middle. Second, in the circular DNA, where the cohesive ends are ligated together, the 1.0- and 4.1-kb fragments are joined, showing that they are at the ends. (b) The P4 phage attachment site in the phage DNA lies in either the 1.0- or 4.1-kb BamHI fragment and not in the 6.4-kb fragment, which is preserved after integration of prophage. The BamHI sites in the host DNA that are nearest the prophage are placed to give host-phage fragments of 15.0 and 12.5 kb.

21 The fragments have eliminated either the promoters or the ribosome binding sites of both genes. The plasmid contains both of these, but they are on the same strand of the plasmid DNA. If genes *A* and *B* are transcribed in opposite directions, only one can be expressed from the plasmid.

22 The reaction has not been carried out long enough and the cuts at positions 35

and 83 were not made.

23 The terminal fragments are missing and a new fragment is present whose size is the sum of the sizes of the missing fragments. Thus, the DNA is circular.

24 Select an enzyme whose restriction site is further from the *lac* promoter than the BamHI site. Let us assume it is HaeI. If the gene of interest does not contain a HaeI site, do the following. Cleave the plasmid with both BamHI and HaeI and retain the fragment (there will be two) that contains the replication origin and other essential genes. One end of the fragment will have a BamHI terminus, and the other end will have a HaeI terminus. Cleave the donor DNA with both enzymes and join the fragments to the isolated plasmid fragment. Since each fragment has one BamHI terminus and one HaeI terminus, joining can only occur in a particular orientation.

25 The principle is to design the system so that expression of the gene can be controlled. For example, a restriction enzyme could be used that separates the coding sequence and the promoter and the coding sequence could be linked to the promoter of the *lac* operon. Then, transcription could be turned on at will merely by adding an inducer of the *lac* operon.

26 $f = 2 \times 10^4/3 \times 10^9 = 6.67 \times 10^{-6}$. Therefore, $N = \ln(0.01)/\ln(1 - 6.67 \times 10^{-6}) = 6.91 \times 10^5$.

MORE CHALLENGING PROBLEMS

1 Plasmid pBR607 DNA (molecular weight, $2.6 \times 10)^6$ carries two genes whose products confer resistance to tetracycline (Tet-r) and ampicillin (Amp-r) in host bacteria. The DNA has a single site for each of the following restriction enzymes: EcoRI, BamHI, HindIII, PstI, and SalI. Cloning into the EcoRI site does not affect resistance to either drug. Cloning into the BamHI, HindIII and SalI sites abolishes tetracycline-resistance and into the PstI site abolishes ampicillin-resistance. Digestion with the following mixtures of restriction enzymes yields fragments with the following sizes:

Enzymes in mixture	*Molecular weights of fragments*
EcoRI, PstI	0.46, 2.14
EcoRI, BamHI	0.2, 2.4
EcoRI, HindIII	0.05, 2.55
EcoRI, SalI	0.55, 2.05
EcoRI, BamHI, PstI	0.2, 0.46. 1.94

Position the PstI, BamHI, HindIII and SalI cleavage sites on a restriction map, relative to the EcoRI cleavage site.

2 The DNA of the plasmid pHUB1 is circular, double-stranded, and contains 5.7 kilobase pairs (5.7 kb). This plasmid carries a gene whose protein product confers resistance to tetracycline (Tet-r) on the host bacterium. The DNA has one site each for the following restriction enzymes: EcoRI, HpaI, BamHI, PstI, SalI, and BglII. Cloning into the BamHI and SalI sites abolishes tetracycline-resistance; cloning into the other sites does not. Digestion with various restriction enzymes and combinations of enzymes yields DNA fragments whose sizes in kilobase pairs are shown below:

Enzymes in mixture	*Fragment size (kb)*
EcoRI	5.7
EcoRI, BamHI	0.4, 5.3
EcoRI, HpaI	0.5, 5.2
EcoRI, SalI	0.7, 5.0
EcoRI, BglII	1.1, 4.6
EcoRI, PstI	2.4, 3.3
PstI, BglII	1.3, 4.4
BglII, HpaI	0.6, 5.1

Position the cleavage sites for all these enzymes on a map of the pHUB1 DNA. Draw the map as a circle with 5.7 kb marked off on the circumference.

3 In analyzing DNA by treatment with a restriction endonuclease followed by gel electrophoresis, in a particular experiment it is observed that in addition to the expected bands, DNA is present in all regions of the gel starting from the position of the most slowly-moving fragment and past that of the most rapidly-moving fragment. If the DNA is analyzed before treatment with the restriction enzyme, no such material is seen. Suggest an explanation.

4 Can the c-DNA technique be used with any eukaryotic gene? This is a practical and not a theoretical question.

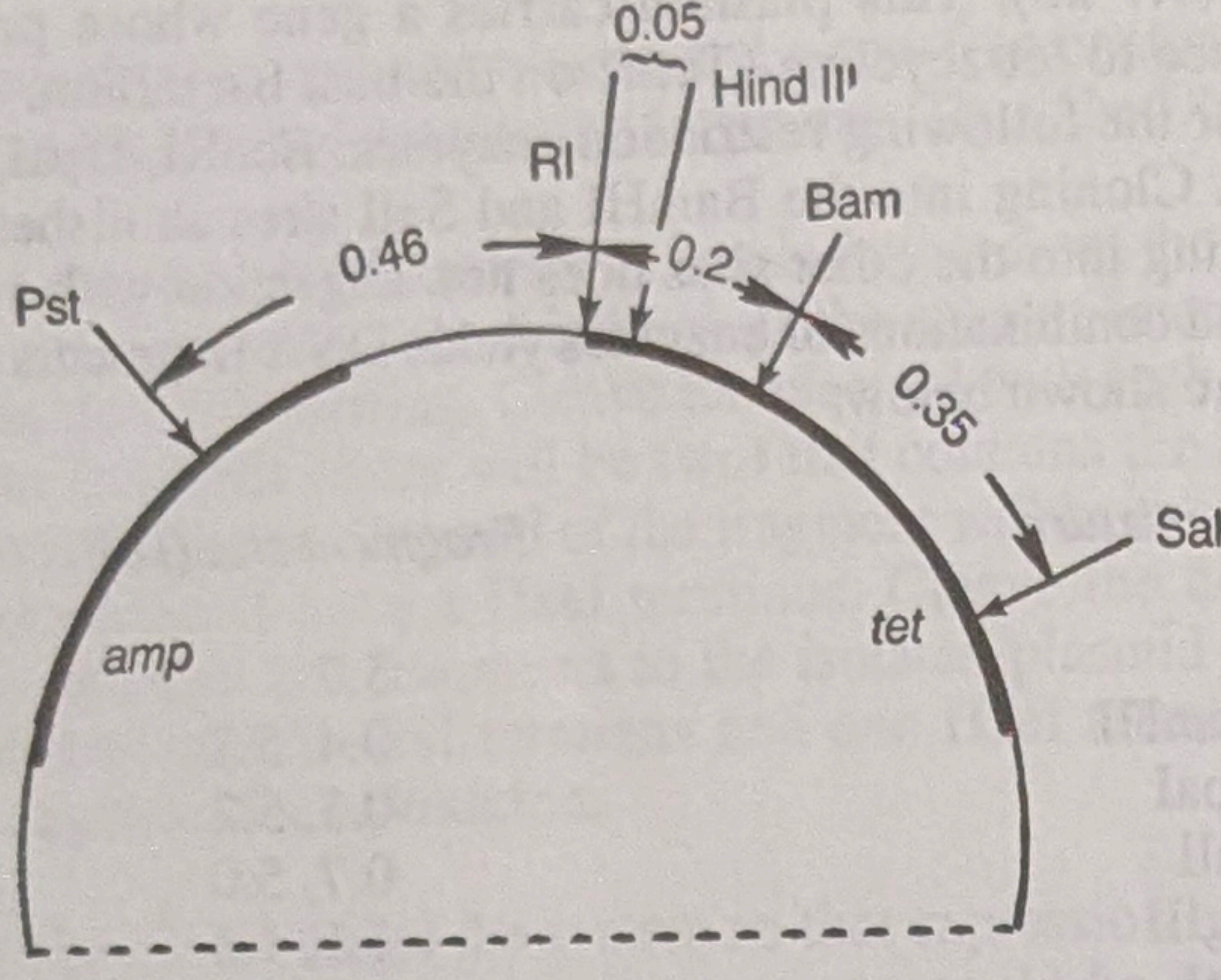
ANSWERS TO MORE CHALLENGING PROBLEMS

For Problems 1 and 2, please see following page

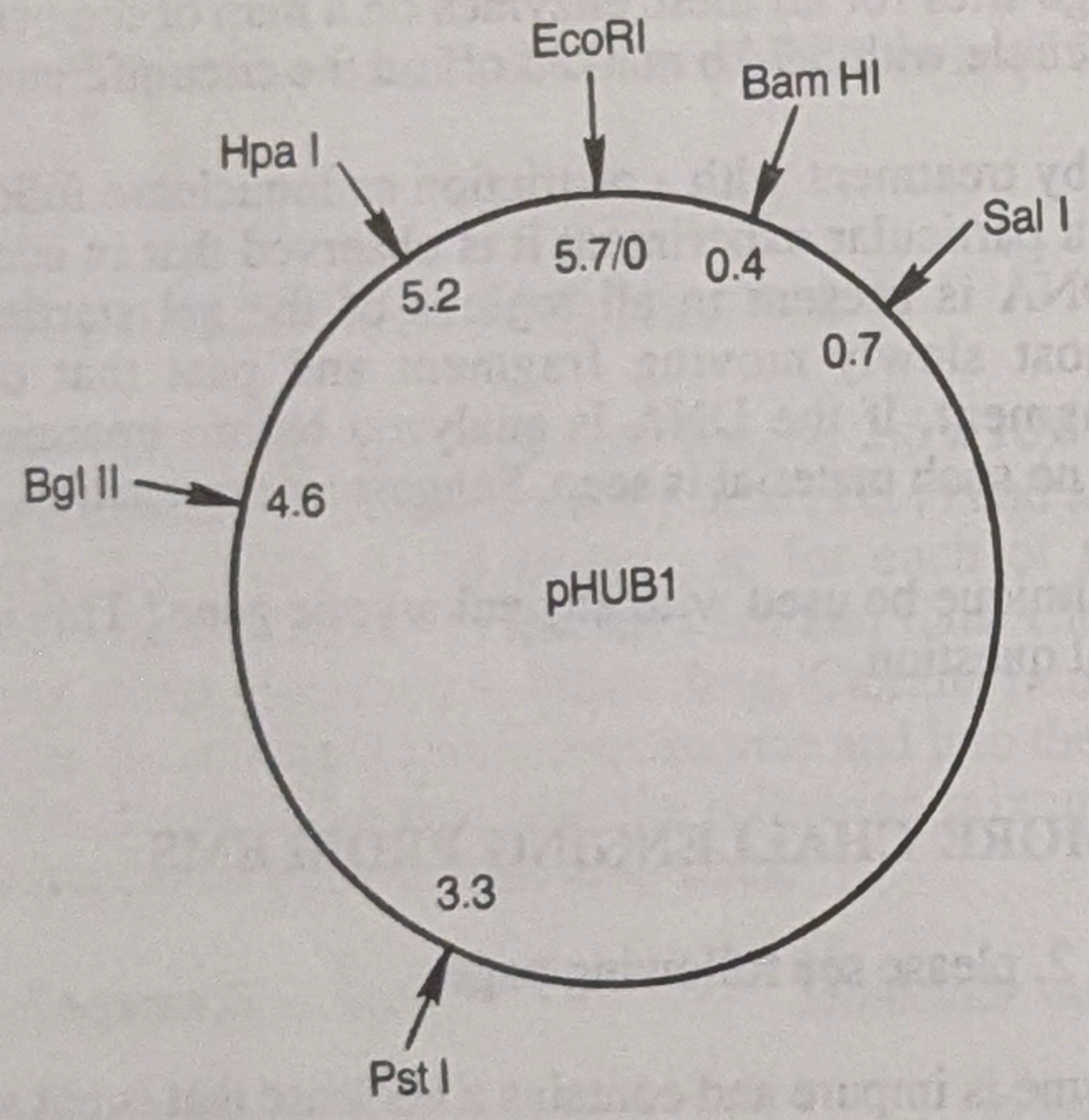
3 The restriction enzyme is impure and contains a nuclease that is not site-specific.

4 No, as a source of fairly pure mRNA is needed. This is only practical with mRNAs made in very large quantities by particular cells, such as the cells making hemoglobin. If the mRNA is a minor species, there might be no way to isolate or identify it. Normally, one would do this by hybridization to a DNA molecule. However, often there is no way to obtain the DNA molecule except by cloning, and the c-DNA method may be the only way to do the cloning.

1



2



CHAPTER 23

EUKARYOTIC VIRUSES

CHAPTER SUMMARY

A variety of viruses of different sorts infect the higher eukaryotes. They contain single- and double-stranded DNA and single- and double-stranded RNA. The DNA may be linear, circular, or linear with hairpin termini. The RNA is linear. Most viruses contain a single nucleic acid molecule, but some, those with segmented genomes, contain more than one; the fragments may all be in a single virus particle (isocapsidic viruses) or in several distinct particles (heterocapsidic particles). The single-stranded RNA molecules in viruses are of two types: (+)-strand, which serves as a messenger RNA, and (–)-strand, which is a noncoding strand.

There are four morphological classes of virus particles: naked icosahedral, enveloped icosahedral, and naked and enveloped helical. The nucleic acid of naked viruses is enclosed only in a protein capsid; the enveloped viruses are further enclosed in a lipoprotein membrane, usually obtained from the cell membrane of the host.

The basic life cycle of the eukaryotic viruses is in principle no different from a phage life cycle. The stages are: adsorption to a host cell, entry of the nucleic acid into the cell, transcription, replication, maturation of particles, and release of particles from the cells. However, the details of the life cycle of a virus are quite different from that of a phage. For example, the entire virus particle usually penetrates the host cell after which an uncoating process occurs, in which the membranes and proteins are removed. This poorly understood process occurs in a variety of ways. Release of progeny virus particles may occur by a lytic mechanism, but not commonly. Some particles are formed in vacuoles, and the vacuole migrates to the cell surface and spills its contents to the surrounding medium. The enveloped viruses are released by budding; the particles form under the cell membrane and an outpocketing of the membrane releases each particle separately. The host cell normally survives both release from vacuoles and budding. Some viruses carry out their life cycle in the nucleus; others are cytoplasmic viruses. The DNA viruses are typically nuclear viruses, since the initial transcription event utilizes host RNA polymerases, which have a nuclear location. The entire life cycle of a typical eukaryotic virus takes 6–48 hours compared to one hour or less for phages.

The animal RNA viruses can be distinguished by their mode of production of mRNA. For example, the virion RNA of the (+)-strand RNA viruses is an mRNA molecule and be can be translated directly. However, the RNA of the (–)-strand viruses is not untranslatable and a (+) strand must first be made. The double-stranded RNA molecules contain an mRNA strand, but translation is not possible from double-stranded RNA; hence, an intermediate step is required. The DNA viruses always have available host RNA polymerase to initiate transcription. However, no host enzyme is able to copy the RNA of a (–)-strand RNA virus or to use double-stranded RNA as a template in the case of the viruses containing double-stranded RNA. Hence, these classes of viruses carry the necessary enzyme in the virion, and the enzyme enters the host cell with the RNA. The RNA of the (+)-strand viruses is not usually used for translation of all proteins. For example, in the case of poliovirus the (+) strand is translated and an RNA replicase is made; this enzyme makes (–) strands, which serve as a template for synthesis of the major mRNA. The retroviruses use their (+) strand as a template for synthesis of double-stranded DNA; the enzyme that carries out this conversion, reverse transcriptase, is also carried in the virion. The double-stranded-DNA intermediate is inserted into the host DNA, and there it is transcribed by the host RNA polymerase.

Since only a single protein can be translated from a eukaryotic mRNA molecule, a variety of mechanisms have evolved that enable the virus to synthesize all of its required proteins. The (–) strand of vesicular stomatitis virus serves as a template for the synthesis of several monocistronic mRNAs. In contrast, poliovirus translates from its RNA molecule a polyprotein, which is subsequently cleaved to form the individual proteins. Reovirus contains segmented double-stranded RNA; the (–) strand of each double-stranded segment serves as a template for synthesis of a (+) strand, each of which is an mRNA for one viral protein. Transcription of inserted retroviral DNA produces several mRNA molecules, some of which form a single viral protein and others a polyprotein that is subsequently cleaved. The double-stranded DNA viruses produce some primary transcripts that encode a single protein, though removal of introns is often necessary. Some viruses, of which adenovirus is an important example, produce a large number of mRNA molecules by numerous alternate modes of splicing of transcripts that start at a small number of sites and terminate at a variety of polyadenylation sites.

Production of progeny viral nucleic acid molecules also occurs by numerous mechanisms. Some DNA viruses replicate in a fairly standard way. The papovaviruses, which contain a circular DNA molecule, replicate bidirectionally as a circle using the θ mode. A specific initiation protein, called T antigen, is needed to start a round of replication. Adenovirus contains a linear DNA molecule with terminal inverted repeats and a protein (the 55-kd protein) covalently linked to each 5' terminus. Replication probably begins by the

binding to the 55-kd protein of an 80-kd protein molecule (the precursor of the 55-kd protein) to which is linked dCTP. The 3'-OH group of this nucleotide is the primer for ongoing synthesis. This initiation event occurs at both 5' ends, so adenoviral DNA replicates from both ends simultaneously. Late in the round of replication the 80-kd protein of the growing strand is cleaved to produce 55-kd protein. The single-stranded-RNA viruses pass through a double-stranded RNA intermediate from which progeny single strands are usually obtained by displacement by growing strands. The retroviruses synthesize their (+) strand by transcription of inserted viral DNA. Synthesis of the double-stranded intermediate is a complex process. In the course of duplication of the RNA, whose termini have sequence 5'–r–u5 and 3'–u3–r, a longer DNA molecule with identical termini having the structure U3–R–U5 (the long terminal repeat or LTR) is formed. Circularization and insertion occur in a poorly understood way

BOLD TERMS

budding, butterfly molecule, capsomers, coviruses, heterocapsidic, hexon, isocapsidic, LTR, minichromosomes, multiple-component viruses, nucleocapsid, penton, polyprotein, protomer, provirus, reverse transcriptase, Rous sarcoma virus, RSV, segmented genome, uncoating, virion

ADDITIONAL INFORMATION

In performing any experiment it is desirable to know the ratio of viruses and cells—that is, the multiplicity of infection. Viruses can be counted by electron microscopy, although this has two disadvantages: (1) for technical reasons the measured value can be as much as three times too great or too small, and (2) there is no way of knowing what fraction of the particles are infective. For this reason assays similar to plaque formation by phages have been developed. A few of these assays are described here.

Many animal cells can be grown as a monolayer in a culture dish. When growth is complete, a confluent layer of cells results which can be stained with various dyes to yield a colored layer. Shortly after the cells are added to the dish, viruses can be added, and some of these will infect cells in the growing layer. If the virus is a type that either causes cell death (for example, poliovirus), or at least significant inhibition of cell growth, there will be a region in the layer in which there are few or no cells when the monolayer has stopped growing. Thus there will be colorless circular regions, called plaques, in the stained background. In contrast with phages for which plaque formation takes 3–12 hours, formation of a viral plaque requires 3–21 days depending on the particular virus, because of the slow growth of animal cells. Some viruses fail to form plaques because they do not kill cells fast enough. However, if a virus (for example, influenza virus) stimulates hemagglutinin synthesis, infectious particles can be assayed. Hemagglutinins are glycoproteins formed on the surface of infected cells; they cause an infected cell to bind a red blood cell. Thus, if a confluent layer of cells containing foci of infection, which are not visible as plaques, is overlaid with a suspension of red cells, and then washed, the foci can be identified as macroscopic red spots on the surface of the monolayer.

Tumor viruses (for example, Rous sarcoma virus) can also be titrated on confluent cell layers. These viruses cause infected cells to continue to grow for several days after the confluent layer has ceased growth. The regions of growth, called foci, appear as raised clusters on the monolayer.

Plant viruses (for example, tobacco mosaic virus) are titrated by abrading the surface of a leaf and then spreading a droplet of a virus suspension on the abraded surface. Dead spots on the leaf, called lesions, are formed; each lesion is the result of infection by a single virus particle.

The efficiency of detection of viruses by each of the bioassays just described is low and varies with the individual virus. When compared with the total number of particles seen by electron microscopy, 0.01–10% of animal virus particles produce a visible infection. (Semliki forest virus and herpes simplex virus are notable exceptions; the best preparations give almost 100% efficiency of infection when assayed under optimal conditions.) The plant viruses are even less efficient. There are three possible causes of the low efficiency of infection—(1) most of the particles may be noninfective owing to defective particle assembly or damage of some sort after assembly, (2) conditions used to detect the virus might, for unknown reasons, limit the ability to form a plaque, focus, or other end point, and (3) successful infection requires the cooperation of many functional particles. It is very important to determine whether alternative 3 is correct if one is to understand the molecular basis of a particular life cycle. In fact, apart from the multicomponent plant viruses in which different particles carry distinct RNA molecules, it is almost always true that successful infection by an animal virus requires only one functional virus particle, as is also true of bacteriophage infection. This is proved by the fact that the number of plaques (or other foci of virus growth) is directly proportional to the number of particles added. The argument for this conclusion can be seen most easily by considering a hypothetical case in which two particles are needed for successful infection.

The Poisson distribution states that if the ratio of the total number of virus particles to cells is m (which in ordinary assay conditions on a petri dish is very small), then the proportion $P(k)$ of cells infected by k particles is $P(k) = e^{-m}m^k/k!$ Thus, if two virions are needed for successful infection, there will be two types of uninfected cells—those infected with no viruses, for which the proportion is e^{-m}, and those infected with one virus, for which the proportion is me^{-m}. Thus, the proportion $P(0)$ of cells that do not produce plaques is $e^{-m} + me^{-m}$ or $e^{-m}(1 + m)$. If m is small, this expression can be approximated as $P(0) = 1 - m^2/2$. Since the proportion of infected cells P^i is $1 - P(0)$, then $P^i = m^2/2$, which means that the number of plaques increases as the square of the number of added virions. It can easily be shown that if more than two virions are required for a successful infection, the number of plaques will increase at a power of m greater than two. On the other hand, if only one virion is needed, the fraction of uninfected cells will be e^{-m} and the fraction of infected cells will be $1 - e^{-m}$. If the number of added virions is increased q-fold, the fraction of infected cells will be $1 - e^{-qm}$. The ratio $(1 - e^{-qm})/(1 - e^{-m})$ is the factor by which the number of plaques increases when q-fold more viruses are added. For small values of m this expression equals q. Thus, when the number of added virions increases q-fold, the number of plaques increases q-fold. Since for small m this conclusion is independent of the value of m—that is, independent of the number of added particles—a linear curve is proof that one virion can initiate an infection.

DRILL QUESTIONS

1 What kinds of nucleic acids are found in virus particles?

2 What is a capsid?

3 What is a capsomer?

4 What is a virion?

5 What is the source of the envelope of the enveloped virions?

6 What is a typical time interval between infection of a cell and release of progeny virus particles, and how does this compare to the life cycle of a typical phage?

7 How are naked viruses usually released from host cells?

8 How are enveloped viruses usually released from host cells?

9 What is the defining characteristic of a (+)-strand viral RNA molecule?

10 What enzyme must be contained in the virion of a (–)-strand RNA virus?

11 What enzyme must be contained in a virion that contains double-stranded RNA?

12 What enzyme, if any, must be contained in a virion with (+)-strand RNA?

13 What kind of nucleic acids are contained in vesicular stomatitis virus, poliovirus, and reovirus?

14 What mechanisms are used for the production of the variety of proteins needed by a eukaryotic virus?

15 What distinguishes naked from enveloped viruses?

16 What are the common basic symmetries of naked viruses?

17 What is meant by the terms "early" and "late" when applied to virus-encoded functions or proteins?

18 What kinds of functions are frequently found to be early?

19 What kinds are late?

20 What is the defining feature of a retrovirus?

21 Is SV40 a nuclear virus or a cytoplasmic virus? Why must this be so?

22 Is reverse transcriptase a viral product or a host product, and where is the enzyme located?

23 What is the template for the production of retroviral mRNA?

24 Do retroviruses replicate in the nucleus or in the cytoplasm?

25 What are the three biochemical activities of reverse transcriptase?

26 What phenomenon suggests that the DNA-insertion mechanism used by retroviruses is related to transposition?

27 What kind of nucleic acid is found in papovaviruses and what is its form?

28 What is the form of deproteinized DNA from papovaviruses?

29 What role do viral T antigens serve in DNA replication?

30 What is the primer for replication of adenoviral DNA?

31 What mechanism is used to regulate the timing of synthesis of adenovirus proteins?

ANSWERS TO DRILL QUESTIONS

1 Single- and double-stranded DNA, and single- and double-stranded RNA. The DNA can be linear or circular.

2 A capsid is the closed protein structure that isolates viral nucleic acid from the environment. It is constructed of protomers.

3 A capsomer is an organized aggregate of protomers. If it consists of five or six protomers, it is a penton or hexon, respectively.

4 A virion is a complete virus particle. It may consist simply of a capsid containing nucleic acids, or the capsid may be enclosed in membranous material.

5 Usually the cell membrane.

6 6–48 hours. For phages, 25–60 minutes is typical.

7 By lysis and by release from vacuoles.

8 By budding.

9 It can either serve as mRNA or it contains the same sequences as viral mRNA.

10 An RNA replicase that converts the (–) strand to either a (+) strand or double-stranded RNA.

11 An RNA replicase that can use double-stranded RNA as a template for synthesis of mRNA.

12 None is required except for the retroviruses, which carry their own reverse transcriptase.

13 VSV and poliovirus contain single-stranded RNA; reovirus has segmented double-stranded RNA.

14 Synthesis of distinct primary transcripts, alternative processing of primary transcripts, and synthesis of polyproteins.

15 Enveloped viruses are encased in membranes.

16 Icosahedral and helical.

17 Early and late refer to time sequence of gene expression in viral development.

18 Viral DNA replication; turning various host cell functions off or on.

19 Synthesis of virion proteins.

20 A DNA intermediate is made from a virion RNA template.

21 It is a nuclear virus since both the host DNA replication enzymes and the host RNA polymerases are contained only in the nucleus.

22 It is encoded in a viral gene. The enzyme is a component of the virus particle.

23 The DNA that is formed by reverse transcription of retroviral RNA.

24 Formation of DNA by reverse transcriptase occurs in the cytoplasm. However, progeny viral RNA is formed by transcription of inserted DNA and hence must occur in the nucleus, where RNA polymerase is located.

25 Copying of single-stranded RNA to form a DNA-RNA hybrid; nucleolytic removal of RNA from a DNA-RNA hybrid; primed copying of single-stranded DNA to form double-stranded DNA.

26 Integrated proviral DNA is terminated by two LTRs in direct repeat and a short sequence in inverted repeat (like a composite transposon) and flanked by a short sequence in direct repeat (like a target sequence).

27 Circular DNA to which histones are bound. These are called minichromosomes.

28 Supercoiled.

29 T antigen is required to initiate DNA replication.

30 A protein to which dCMP is linked.

31 Different patterns of RNA processing occur at different times and these produce distinct mRNAs.

PROBLEMS

1 What kinds of viruses must carry enzymes for mRNA synthesis as part of the virion and introduce them into the cell upon infection?

2 What kinds of viruses must carry enzymes for DNA synthesis as part of the virion and introduce them into the cell upon infection?

3 Some viruses contain substantial quantities of phospholipids, cholesterol, and carbohydrates. What is the source of these substances?

4 What features do (+)-strand viruses share with eukaryotic mRNA?

5 With any single-stranded RNA virus, which strand has a ribosomal binding site?

6 With phages, the efficiency of plating (EOP, the fraction of particles that can form a plaque) is generally nearly 1; with viruses, values of from 10^{-6} to 10^{-1} are more common. What factors contribute to the low values of EOP of viruses?

7 RNA viruses have four distinct mechanisms for producing mRNA. What are these mechanisms?

8 Viruses must synthesize several proteins and must deal with the problem that in eukaryotes only a single protein can be translated from an mRNA molecule. What mechanisms are used to circumvent their problem?

9 The total molecular weight of several animal viruses exceeds the coding capacity of the virus, if a single reading frame is used. Three major mechanisms contribute to this phenomenon. What are they?

10 A pulse-chase experiment in virus-infected cells is performed with ^{3}H-leucine. At various times the molecular weights of the labeled proteins are determined using SDS-PAGE electrophoresis, which employs conditions that denature the proteins and dissociate subunits. What can you conclude about particular viral systems in which the following observations are made? (a) If the labeling is done between 2 and 3 hours after infection, label is present at 4 hours but not at 12 hours? (b) One labeled band is found 3 hours later and 6 bands are found 8 hours later. The molecular weights of the 6 proteins are smaller than those of the proteins at 3 hours postinfection. (c) One labeled band (molecular weight = 6×10^5) is seen at early times, 3 (molecular weights = 5.6×10^5, 25×10^3, and 15×10^3) at a later time, and 7 (molecular weights = 11×10^4, 8×10^4, 7×10^4, 6×10^4 5×10^6, 4×10^6, and 3×10^6) at very late times.

11 What mechanism ensures that each reovirus particle gets one copy of each RNA molecule of its segmented genome?

12 How do VSV and poliovirus differ with respect to the mechanism for producing the various viral proteins?

13 What is the difference between the way that integration fits into the life cycles of retroviruses and the life cycles of the DNA tumor viruses?

14 Certain features of the life cycles of viruses are analogous to those in phage-infected bacteria. Examples are establishment of new replication and transcription systems, synthesis of huge amounts of structural proteins that self-assemble, and except for the simplest viruses, and distinct periods of early and late transcription. However, certain features of viral systems are rarely found, if ever, in phage systems. Name some of these features.

15 What forms of nucleic acid are isolated from polyoma virus and from SV40?

16 What mechanism does SV40 utilize that allows the total molecular weight of its protein to exceed the coding capacity of its DNA?

17 Adenovirus DNA molecules are able to circularize yet they do not have comp-

lementary single-stranded termini. How do they circularize?

18 What is the meaning of the terms "permissive" and "non-permissive" in viral infections?

ANSWERS TO PROBLEMS

1 Any kind that contains genetic material that is incapable of directing cellular enzymes to make mRNA. For example, (–)-strand RNA viruses, like influenza virus; double-stranded RNA viruses, like reovirus; double-stranded DNA viruses that multiply outside the nucleus, like vaccinia; and RNA viruses that replicate through a DNA intermediate, like retroviruses.

2 Retroviruses, which copy DNA from an RNA template, and vaccinia, which replicates its DNA in the cytoplasm where there are no DNA polymerases.

3 Host cell outer membranes.

4 Poly(A) tail, and, except for poliovirus RNA, a cap.

5 The (+) strand.

6 Noninfective particles, experimental conditions that limit detectibility, and a requirement for the participation of more than one virion for successful infection.

7 Double-stranded RNA viruses carry into the cell a replicase that copies double-stranded RNA and synthesizes a (+) strand; (–)-strand viruses carry into the cell a replicase that synthesizes mRNA from the (–)-strand template; (+)-strand RNA (other than retrovirus RNA) is translated directly; retrovirus RNA is used as a template for synthesis of DNA, which is then transcribed.

8 The (+)-strand viruses, such as poliovirus, synthesize a polyprotein that is cleaved to form individual proteins. Some (+)-strand viruses also make a (–) strand from which short (+) strands are made. The (–)-strand viruses, of course, first make a (+) strand; in some viruses this (+) strand is cleaved. In other viruses, a polyprotein is made, which is then cleaved; still others process the (+) strand and cleave polyproteins.

9 Different cleavages of polyproteins; different cleavage patterns of mRNA, producing start codons in different reading frames; different splicing patterns of mRNA.

10 (a) The proteins labeled between 2 and 3 hours are degraded. (b) A polyprotein is made at early times and it is degraded in at least two steps. (c) A polyprotein is made that is first cleaved into three components. Since the proteins whose molecular weights are 25×10^3 and 15×10^3 are not present at late times, they must be unstable. The 5.6×10^5-protein is cleaved; since the sum of the molecular weights at late times is less than 5.6×10^5, either there are several copies of (at least) one cleavage product or one or more of the products is unstable.

11 Details are unknown, but either each replica remains associated with a replica of the core protein or, more likely, specific pre-core proteins bind to particular RNA molecules (base-sequence-specific binding) and these pre-core proteins aggregate specifically to form a core.

12 VSV makes several mRNA molecules, each of which encodes one protein. Poliovirus makes a polyprotein, which is cleaved.

13 Integration is obligatory in the life cycle of a retrovirus.

14 Entry of viral proteins into the infected cell, processing of mRNA, production of polyproteins, conversion of RNA to DNA.

15 Supercoiled DNA. Occasionally, if the DNA is not isolated carefully, open circles and even linear molecules can be found. If the DNA is not subjected to procedures that remove proteins, all histones other than H1 are associated with the DNA, which has a nucleosome-like form.

16 Differential splicing allows different reading frames to be used and removes particular start and stop codons. Some proteins are translated in different reading frames and others are totally contained in the sequence of a larger protein. No proteins are cleaved from polyproteins.

17 The 5' termini of the two single strands of adenovirus double-stranded DNA are covalently linked to a protein molecule that dimerizes.

18 Permissive—progeny virus are produced; nonpermissive—viral nucleic acids enters the host cell but progeny virus are not produced.

CHAPTER 24

TUMOR VIRUSES AND ONCOGENES

CHAPTER SUMMARY

In the course of differentiation of a fertilized egg to form a mammal most cell types lose the ability to divide. In the adult organism only a few classes of cells divide. Many cells retain the potential to divide, and, following an injury, healing takes place by cell division. Occasionally, cells whose growth is normally kept in check by unknown factors begin to divide, and a tumor develops. If the cells of the tumor adhere poorly to one another, individual cells can leave the tumor and establish clones elsewhere in the body; these cells are malignant or cancerous. The general lack of ability of cells to divide indefinitely is manifest in cell culture: cells taken from animal tissue and placed in culture medium divide only a limited number of times and then, if kept in culture, they die and disintegrate. This stage of a cell culture is called crisis. If the culture is maintained for many months, a few rare cells which presumably have suffered some mutation, begin to divide and form clones that are capable of indefinite growth. These cells constitute an established cell line; since they have gained the ability to divide an unlimited number of times, the culture is said to be immortalized. Cells in established cell lines retain certain growth characteristics of cells in tissue; for example, they multiply only until all cells are in contact, forming a monolayer on the surface of the vessel in which they are growing. This phenomenon is called density-dependent growth or contact inhibition. If established cell lines are transplanted to an animal, in general, tumors will not develop.

If cells of a cell line are exposed to carcinogens or infected with tumor viruses, progeny cells are produced that have lost certain features of growth control. These cells, which are said to be transformed, have the following properties: growth to higher cell densities than the original cell line (they do not exhibit contact inhibition), a lesser requirement for certain growth factors that are normally added to the culture medium, poor adhesion to one another, growth without being anchored to a solid surface, more fluid cell membrane, chromosome number exceeding the diploid number, and, significantly, the ability to cause tumors when injected into animals. Most, if not all, of these changes, are a result of the activity of particular genes called oncogenes.

The presence of oncogenes in cancer cells can be demonstrated by a simple test. Cells of a nontransformed cell line, usually, NIH 3T3 mouse cells

ability to cause tumors when injected into animals. Most, if not all, of these changes, are a result of the activity of particular genes called oncogenes.

The presence of oncogenes in cancer cells can be demonstrated by a simple test. Cells of a nontransformed cell line, usually, NIH 3T3 mouse cells,

are exposed to DNA isolated from cancerous tissue. Some cells in the mouse cells become transformed. Exposure to DNA from normal tissue does not produce transformation. Tests of individual fractions of the sample of DNA from the cancerous cells demonstrates that transformation results from incorporation of a particular nucleotide sequence. These sequences are usually identical to, or slight variants of, sequences present in normal cells. Such a sequence in a normal cell is called a protooncogene. In the first such study, DNA from human bladder cancer was found to contain an oncogene that differed from the protooncogene by only a single base pair; that is, the oncogene, which was termed *ras,* was a mutated protooncogene. The *ras* oncogene has turned up in several human cancers. A few experiments indicate that, at least in some cases, carcinogens, which are mutagenic, convert protooncogenes to oncogenes by introducing a single-base mutation into the protooncogene.

Many viruses cause cancer. These viruses all have the ability to insert their DNA, or in the case of retroviruses, the DNA copy of the virion RNA, into the chromosome of the animal. The DNA tumor viruses, which insert their DNA only infrequently, only cause cancer if an animal is infected with a very large number of viruses. The retroviruses, for which integration of viral DNA is an integral part of their life cycle, are the most tumorigenic viruses known. A variety of viral oncogenes have been detected in the RNA tumor viruses. In each case studied the oncogene is not essential for production of progeny virus and is a copy or a mutated form of a host protooncogene. To distinguish the host and viral forms, the term *c-onc* is used for the protooncogene and *v-onc* for the viral oncogene. The *c-onc* sequences generally contain introns, whereas the *v-onc* sequences do not. Presumably the viral gene was acquired, in ancient times, from a c-DNA copy of a processed *c-onc* mRNA.

The DNA tumor viruses may not always carry oncogenes and transformation by those that do not may result from placing a host gene under control of a viral promoter or virus-induced chromosome breakage and rearrangement that leads to changes in the expression of critical growth-regulating genes.

Rous sarcoma virus contains an oncogene called *v-src*. This oncogene encodes a tyrosine protein kinase, an enzyme that phosphorylates tyrosine moieties in a protein. Similar oncogenes have been found in other retroviruses. Protein kinases are common in cells but they are rarely active against tyrosine. The significance of the kinases is that they can add phosphate groups to a large number of cellular proteins; thus, there is no difficulty imagining that a protein kinase could activate a positive effector or inactivate an inhibitor of cell division. About 30 oncogenes have been isolated from various cancers and tumor viruses. Many oncogenes are common to different cells and different viruses. These genes encode a variety of proteins, for example, threonine kinases, nuclear DNA-binding proteins, a protein that resembles a growth factor, an altered form of the receptor for another growth factor, and proteins that bind

GTP. It is not known how any active oncogene removes growth control.

BOLD TERMS

cell fusion, contact inhibition, crisis, density-dependent growth, established cell line, growth factors, heterokaryon, initiation, oncogene, oncogenic, primary culture, promotion, protein kinase, protooncogene, transformed, transformed cell, translocation, tyrosine-specific protein kinase

ADDITIONAL INFORMATION

One approach to understanding the role of oncogenes in tumor cells and of the protooncogenes in normal cells is to neutralize the gene products with an antibody. Such experiments have been done recently with a pure antibody (a monoclonal antibody) prepared against *ras* proteins. This antibody is active against both *v-ras* and *c-ras* proteins. Microinjection of the antibody into nontransformed NIH 3T3 cells causes an inhibition of the initiation of DNA synthesis 12–15 hr after injection, with no effect on DNA synthesis already in progress. Thus, it is possible that *v-ras* causes cellular transformation in NIH 3T3 cells simply by allowing DNA synthesis when it should not occur. It is conceivable that either an excess amount of *ras* protein or a mutant *ras* (which might not respond to some inhibitory signal) is sufficient to overcome some normal (and unknown) block to initiation of synthesis. A suggestion that this may be part of the story comes from recent experiments in which cells stably transformed by *ras* DNA were microinjected with anti-*ras* antibody. Initiation of DNA synthesis in these cells occurred, as expected, but interestingly the morphology of the cells reverted from the rounded appearance of transformed cells to the spindle shape of the normal NIH 3T3 cells. Cells transformed by several other oncogenes were also studied, and reversion to normal morphology occurred with cells transformed with the growth-factor receptorlike oncogenes (*src, fems, fes*), but not with other oncogenes (*mos, raf*). While the interpretation of these experiments is complex and many more experiments are necessary, it is clear that transformation by some oncogenes requires a functional normal amount of intracellular *ras* protein. For additional information, refer to Smith et al, *Nature, 320,* 540, 1986.

An article, which might be of interest, since it is written by a physician who views protooncogenes as normal cellular proteins involved in mitosis, is: H. Gordon, *Mayo Clinic Proc., 60,* 697, 1985. This article is especially readable and summarizes a great deal of what was known in 1985 about oncogenes.

DRILL QUESTIONS

1 What is the main feature that distinguishes an established cell line from cells just obtained from tissue or from a primary culture?

2 What three features of the growth of transformed cells distinguish them from a nontransformed cell line?

3 What is the major reason for the greater oncogenicity of retroviruses compared to the DNA tumor viruses?

4 If cellular DNA from a virus-transformed cell was treated with a restriction enzyme that makes no cuts in the viral DNA, how would one distinguish lack of integration of the viral DNA from integration of a single copy of the viral DNA?

5 If cellular DNA from a virus-transformed cell was treated with a restriction enzyme that makes a single cut in the viral DNA, how would one distinguish lack of integration of the viral DNA from integration of a single copy of the viral DNA? Assume that virion DNA is circular.

6 A human cell is infected with a particular tumor virus. Five different virus-transformed clones are isolated and studied by the cell fusion technique. Every mouse-human hybrid cell line containing human chromosome 7 shows the transformed phenotype. This is not true of any other human chromosome. What can be said about transformation by this virus?

7 Tumor viruses have been likened to lysogenic phages because of the integration of their DNA. Comment on the differences between the transcription of prophages and proviruses.

8 In what way does the location of a provirus differ from the location of a prophage?

9 Can a Rous sarcoma virus in which the *src* gene has been deleted or partially deleted replicate?

10 Do *c-onc* or *v-onc* genes contain introns?

11 What biochemical reaction does a protein kinase carry out?

12 There are many protein kinases, but what is unusual about the protein kinases encoded in several oncogenes?

13 A particular oncogene protein has the same amino acid sequence as that of a protooncogene. In what ways could the presence of the gene lead to transformation?

14 In a given animal type is each kind of cancer (that is, cancer in different tissues)

transformation occurs?

5 How could chromosome breakage and rearrangement lead to activation of an oncogene?

6 An oncogene makes a tyrosine kinase that differs from that of the protooncogene by a single amino acid. What features might the altered form have that would make it oncogenic?

7 Certain growth factors are essential for multiplication in culture of some established cell lines. Transformed variants of these cells sometimes do not require these factors any longer. Suggest some possible explanations.

8 If a cell with 22 chromosomes is fused to another species of cells with 30 chromosomes, what is the maximum and minimum number of chromosomes found in cells that are hybrid for some particular gene?

9 How could one prove by the cell-fusion technique that a given polyoma-transformed cell line contains the entire viral genetic material? Complete base-sequencing should not be considered an option.

10 What kind of viral mutants of a DNA-containing tumor virus might be able to transform cells that are permissive for the wild-type virus?

11 A cell line is transformed with virus PY78. Ten transformed cell lines are isolated. The DNA from each transformed line is isolated and treated with a restriction enzyme that does not make cuts in viral DNA. After gel electrophoresis and Southern blotting with labeled viral DNA as probe, each of the 10 cell lines is found to have a single radioactive band. The band is in the same position in all cell lines. Use of a different restriction enzyme, which also does not attack viral DNA, yields a single band for each cell line. The band is not in the same position as that obtained with the first enzyme. What do these results tell you about the state of the viral DNA in the cell? That is, where is the viral DNA and how many copies are present?

12 A cell line is transformed with virus Q. The DNA of the transformed cell is isolated and degraded with a restriction enzyme known to make a single cut in the viral DNA. After gel electrophoresis and Southern blotting with labeled viral DNA as a probe, four bands are observed. None have the size of viral DNA, and each has the same density on autoradiographic film. What does this tell you about the viral DNA in the transformed cell?

13 A cell culture is transformed with virus RX3. Two transformed cell lines, R1 and R2, are isolated and tested as in Problem 11 using a restriction enzyme that makes a single cut in the viral DNA. Gel electrophoresis of the treated DNA from cell line R1 yields 3 bands, one of which migrates at the same rate as linear

viral DNA. The intensities of the 3 bands are identical. The cell line R2 also yields 3 bands, one of which has the size of linear viral DNA; the intensity of this band is twice that of each of the other bands. To be sure that the bands with virus-sized DNA represent a viral genome rather than a mixed virus-host fragment whose size just happens to be that of viral DNA, the DNA of each cell line is treated with a second restriction enzyme that also makes a single cut in viral DNA. Again viral-DNA-sized bands are observed with the same intensity relations described for the other enzyme. What is the state of the viral DNA in these two cell lines?

ANSWERS TO PROBLEMS

1 They are cancer cells obtained from tumor tissue.

2 If all human chromosomes have been studied, the integrated viral DNA must be in human chromosome 3. However, there are at least eight distinct sites of integration within that chromosome.

3 No. In this experiment a selection has been made for transformed cells. It is possible that insertion can occur at many sites but only at that site will transformation occur.

4 Transformation by DNA tumor viruses only occurs in hosts that are non-permissive for development of progeny virus. These viruses always have an early and a late phase of transcription. In the nonpermissive hosts late transcription does not occur; the mechanism is not known.

5 It could move it next to a stronger promoter, it could separate it from an inhibitor, or both. More complicated suggestions can also be made.

6 It might be more or less active, it might not respond to some regulatory substance, and it might be able to phosphorylate a particular protein against which the wild-type form is inactive.

7 One possibility is that an alternative pathway for some process is present in the transformed cell, and this pathway bypasses the need for the growth factor. Another possibility that is suggested by the fact that some oncogenes encode growth factors is that the transformed cell synthesizes the growth factor itself. In fact, some transformed cells have been found with this property.

8 The maximum is 22 + 30 = 52. The minimum is 22 + 1 = 23 or 30 + 1 = 31, depending on what kind of cell the hybrid is.

9 Recall that transformation only occurs with nonpermissive cells. Thus, if transformed cells are fused with permissive cells, induction of the production of

viable viruses suggests that the entire genome is present in the transformed cell. Actually, only essential genes are necessary. However, the small DNA viruses do not seem to carry nonessential genes.

10 A mutant that either eliminates a killing function or that cannot carry out late transcription might transform a permissive cell.

11 Integration occurs at a single position in all of the 10 cell lines. Nothing is known about the number of copies because an enzyme that cuts viral DNA was not used.

12 There are two copies of the viral DNA, and they are at different locations.

13 The existence of DNA of virus size implies tandem copies of viral DNA. Thus, line R1 has two viral DNAs in tandem. In line R2 the doubled intensity means that there are twice as many copies of viral DNA as there are viral sequences in any other band. Thus, in line R2 there must be three viral DNA molecules in tandem.

JONES AND BARTLETT PUBLISHERS, INC.

ISBN 0-86720-070-7